

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/19266>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Decision support in the management of Rheumatoid Arthritis

Decision support in the management of Rheumatoid Arthritis

Een wetenschappelijke proeve op het gebied der Medische Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
op gezag van de Rector Magnificus Prof. Dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op donderdag 8 mei 2002,
des namiddags om 13.30 uur precies

door

Jakob Fransen
geboren op 30 juli 1968 te Assen

Promotores: Prof. Dr. P.L.C.M. van Riel, Nijmegen
Prof. Dr. G. Stucki, München
Prof. Dr. B.A. Michel, Zürich

Manuscriptcommissie: Prof. Dr. J.M.W. Hazes, Rotterdam
Prof. Dr. R.P.T.M. Grol, Nijmegen
Prof. Dr. A.L.M. Verbeek, Nijmegen

Financial support:

The Swiss Clinical Quality Management in Rheumatoid Arthritis (SCQM) was sponsored by grants from the Swiss Health Authorities (BAG), the Swiss Academy for Health Sciences (SAMW), and the Böni Foundation. Currently the SCQM is financed through a finance pool, with contributions from Aventis, Essex, Merck Sharp & Dome, Roche, and Wyeth (all Switzerland).

The printing of this thesis was financially supported by the SCQM, Amgen, Aventis, Schering-Plough, and Wyeth (all Netherlands).

Cover design: Jurjen Fransen
Printing: Drukkerij Ipskamp, Enschede

ISBN: 90-9016715-3

Contents

Chapter 1	General introduction
Chapter 2	The merits of monitoring: should we follow all our RA patients in daily clinical practice? <i>Rheumatology-Oxford. 2002;41(6):601-4.</i>
Chapter 3	Clinical quality management in rheumatoid arthritis: putting theory into practice. Swiss Clinical Quality Management in Rheumatoid Arthritis. <i>Rheumatology-Oxford. 2000;39(5):542-9.</i>
Chapter 4	Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. <i>Rheumatology-Oxford. 2000;39(3):321-7.</i>
Chapter 5	Responsiveness of the self-assessed Rheumatoid Arthritis Disease Activity Index (RADAI) to a flare of disease activity. <i>Arthritis Rheum. 2001;44(1):53-60.</i>
Chapter 6	The effectiveness of a measurement-feedback system on outcome in rheumatoid arthritis: a controlled clinical trial. <i>Accepted by the Annals of the Rheumatic Diseases</i>
Chapter 7	Influence of guideline adherence on outcome in a randomized controlled trial on the efficacy of methotrexate with folate supplementation in rheumatoid arthritis. <i>Submitted</i>
Chapter 8	Rheumatologists' opinion on the feasibility of a measurement-feedback system in RA – and the influence of motivation. <i>Rheumatology-Oxford. 2003;42:1-5.</i>
Chapter 9	Design and analysis of a RCT on the effectiveness of clinical decision support on physician performance and health outcomes in the management of RA. <i>Accepted by Arthritis Care and Research</i>
Chapter 10	Summary
Chapter 11	Samenvatting
	Publications
	Dankwoord
	Curriculum Vitae

Chapter 1

General introduction

Rheumatoid Arthritis (RA) is a systemic disease that is mainly characterised by symmetric erosive synovitis and occasional multisystem involvement. Most patients exhibit a chronic fluctuating course of disease that, if left untreated, results in progressive joint destruction, deformity, disability and premature death. It frequently affects patients in their most productive years, and thus, disability results in a major economic loss [1]. The progressive joint destruction is the most prominent feature of RA, and is now regarded as the result from the interaction of synovial hyperplasia, chronic inflammation, and autoimmunity [2,3]. However, the precise etiology of RA is still unclear.

Important developments in the last decade influenced the management of RA. Mainly for biological and immunological reasons, it was recognised that a progressive inflammatory disease as RA should be treated with disease modifying anti-rheumatic drugs (DMARDs) as early as possible after disease onset [4]. The availability of more effective DMARDs, use of combination therapies in additive, saw-tooth or step-down schemes, and the introduction of new “biological” agents opened new treatment options. For the first time, it was possible to reach relatively large treatment effects in time spans of 3-6 months, and even to induce remission [5]. Instruments for assessing process and outcomes of RA were developed, while standardised use for clinical trials, cohort studies and clinical practice was promoted [6-8]. This made it possible to assess the effects of RA therapy more closely, and made outcome assessment in reach of daily clinical practice [9].

In the management of RA, there is now general agreement that rheumatoid inflammation must be controlled as soon as possible, as completely as possible, and that this control should be maintained for as long as possible, consistent with patient safety [5]. Further, more attention is paid to the integration of outcome assessment in the management of RA. To support the management of RA, the Swiss Clinical Quality Management in Rheumatoid Arthritis (SCQM) was established in 1997 [10]. Core activity of the SCQM is to provide rheumatologists with a measurement-feedback system to monitor the course of RA. Using the feedback, drug therapy can be optimised to reduce disease activity, leading to prevention of joint damage and concurring disability.

The main subject of this thesis is to study if measurement-feedback, or decision support, is effective in the management of RA.

Monitoring and documentation of rheumatoid inflammation is not used regularly in daily clinical practice [11-13]. 'Many, if not most, clinicians feel it is not necessary to document obvious clinical improvement recognised by both the patient and the clinician. A positive response to the question "How are you doing?" is sufficient' [14]. In **Chapter 2**, it is argued that, to evaluate if treatment goals are reached and to support adaptation of the treatment program, the management of RA patients in daily clinical practice should include systematic and regular evaluation of rheumatoid inflammation.

The major practical problem in monitoring is: how does a clinician know when rheumatoid inflammation is optimally controlled?; what are the criteria to state that inflammatory activity is under control? The disease activity score (DAS) and the EULAR response criteria are suited to aid in determining and evaluating actual status and change in status [15-17], particularly when applied to individual patients with RA. The reason is, that the response status according to the EULAR criteria is not only dependent on the magnitude of change, as with the ACR response criteria [18], but also depends on the absolute level of inflammatory activity reached. A low level of inflammatory activity over time reduces the probability of progression of radiological visible joint damage [19]. With the help of regular monitoring of the DAS, rheumatoid inflammation can be controlled by "titration" of the DMARD dose. The principle of dose titration was demonstrated in 2 studies in RA, using a step-up and a step-down scheme, respectively [20,21].

In **Chapter 3**, the use of a monitoring system, the Swiss Clinical Quality Management in Rheumatoid Arthritis (SCQM), for daily clinical practice is demonstrated in 4 RA patients. Management aimed at reduction of inflammatory activity and pain in the short-term, and joint damage, and consequently disability, in the long-term.

The advantages of the use of graphical display of clinical status and disease scores for decision support and improving compliance are illustrated. Also, it is demonstrated that there is a need to be cautious when using indices such as the DAS for individual patients. Most measures are developed based on more or less homogeneous groups of patients. A measure is most likely to be valid for an individual, if that patient represents sufficient characteristics of such a group.

As analogue or complement to the assessment of RA disease activity with the DAS, the patient assessed Rheumatoid Arthritis Disease Activity Index (RADAI) has been developed [22]. The aim of the RADAI is to provide an easy to use assessment of RA

disease activity, which serves as a complement to the physician's assessments, with which the physician's assessment could be omitted in certain situations. The RADAI is used in the monitoring system of the SCQM (Chapter 3), and as the primary outcome measure in a controlled clinical trial on the effects of the monitoring system on disease activity (Chapter 6). In **Chapter 4**, it is described how a cross-sectional sample of 584 RA out-patients was used to assess the internal consistency and the convergent validity of the RADAI. It was shown that the single index approach is valid, and that the RADAI as measure of patient perceived disease activity is related with, but may not automatically replace, other measures of disease activity such as the DAS28.

While it was shown that the RADAI is a valid measure of disease activity cross-sectionally, its responsiveness, or ability to capture clinically important changes, was still unknown. Measures that are cross-sectionally valid are not necessarily responsive [23]. In **Chapter 5**, a post-hoc analysis of data from a randomized double blind controlled trial of MTX versus Collagen II [24] provided evidence that the RADAI is sensitive to detect relevant increases in disease activity in patients with RA. The changes in the RADAI were well correlated with changes in the DAS28. The discriminative ability of the RADAI and of the DAS28 to detect a flare, reflected by the area under the ROC curve, were virtually the same, as were the effect sizes. Thus, there is evidence that the RADAI can capture relevant increases in RA disease activity.

The main reason to implement a measurement-feedback system in daily clinical practice is, that it could help to adjust the treatment strategy for individual RA patients for achieving optimal control of disease activity. In **Chapter 6**, a controlled clinical trial with 48 rheumatologists and 264 patients is described, to study whether measurement-feedback is effective. Afterwards, the sample appeared to comprise users and non-users of the feedback system. Use was associated with a reduction of disease activity in the feedback period as compared to the control period, in patients with high disease activity at baseline. However, how far changes in medication strategy played a role remained uncertain. Further, there were no explicit guidelines on the management of RA provided. It is possible that the combination of systematic monitoring with explicit guidelines is an adequate tool to promote changes in medication strategy.

In **Chapter 7**, the influence of guideline adherence on outcome was studied by post-hoc analysis of a 48-week, randomised, double blind and placebo controlled trial, on the

effect of suppletion of folic or folinic acid on toxicity and efficacy of MTX treatment in RA [20]. The MTX dose was steadily increased according to the level of disease activity, using a guideline comparable to the EULAR criteria [16]. However, the rheumatologists were allowed to deviate from the guideline. It appeared that adherence to the guideline for MTX dosing had an influence on drug dose and efficacy. Prescribing a MTX dose lower than the guidelines proposed reduced efficacy, and had no clear beneficial effect on toxicity. The effect of prescribing doses higher than the guidelines proposed is unclear, but it did not seem to be more beneficial. The study was not comparing the use of guidelines versus no guidelines. Thus, it can not be stated that guidelines will be effective in clinical practice. However, as it can be expected that in clinical practice the variation in prescribing MTX dose is much larger, guidelines with appropriate adherence may influence efficacy, and perhaps toxicity, in practice.

For reasons of efficiency, it is important that a measurement-feedback system is feasible in daily clinical practice, and is appreciated by its users. Further, it is useless to study the efficacy of decision support if compliance is low, especially if the study outcome depends heavily on patient health outcomes. In **Chapter 8**, a survey is described, performed to assess rheumatologists' opinion about feasibility of the measurement-feedback system in RA. In addition, it could be analysed whether motivational aspects play a role in perception and use of the measurement-feedback system.

Rheumatologists joining the measurement-feedback system for the evaluation of their individual RA patients ("internal motivation") were more satisfied with the system than rheumatologists joining because of scientific purposes or as obligation ("external motivation"). Further, rheumatologists with "internal motivation" perceived the measurement-feedback system as more useful, were more satisfied with the feedback report, less bothered by the time consumption, and also made more use of it. Most important barriers concerned the practical use for decision-making, lag time of the feedback, and time consumption. Influencing motivation and specific reduction of effort might increase overall acceptance and use of the measurement-feedback system for decision support.

The combination of clinical guidelines and systematic evaluation could be a valuable decision support in optimizing the management of RA patients. However, the efficacy of decision support in the management of RA remains unproven. In **Chapter 9**, a proposal is made for design and analysis of a randomized controlled trial on the efficacy of a

decision support system on physician performance and health outcomes in the management of RA.

A decision support intervention primarily aims at the physician. Therefore, a cluster randomized controlled trial, where physicians are randomized rather than patients, is the most appropriate design. Using the proportion of patients with a low inflammatory activity ($\text{DAS28} \leq 3.2$) as primary outcome, it was estimated that a sample size of 238 RA patients and 22 rheumatologists would be needed, which is 2.5 times larger than when clustering is ignored. Analysis can be performed with Multilevel Analysis, which is a complex technique but has the advantage that it deals with multiple variables and produces estimators for the strength of effects [25]. To prevent bias in this particular design, patient recruitment has to take place before randomisation, the control intervention has to be carefully designed, the primary outcome has to be measured by independent assessors, and drop-outs should be prevented as much as possible.

References

1. Guidelines for the management of rheumatoid arthritis. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. *Arthritis Rheum* 1996;39:713-22.
2. Gay S, Gay RE, Koopman WJ. Molecular and cellular mechanisms of joint destruction in rheumatoid arthritis: two cellular mechanisms explain joint destruction? *Ann Rheum Dis*. 1993;52(Suppl 1):S39-47.
3. Bresnihan B. Pathogenesis of joint damage in rheumatoid arthritis. *J Rheumatol*. 1999;26(4):717-9.
4. Klareskog L, Nordmark B, Lindblad S. On the organization of an early arthritis clinic. *Best Pract Res Clin Rheumatol* 2001;15(1):1-15.
5. Moreland LW, Russell AS, Paulus HE. Management of rheumatoid arthritis: the historical context. *J Rheumatol* 2001;28:1431-52.
6. Boers M, Tugwell P, Felson DT et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol* 1994;21:86-9.
7. Wolfe F, Lassere M, Van der Heijde D, Stucki G, Suarez-Almaroz M, Pincus T et al. Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. *J Rheumatol* 1999;26:484-9.
8. Pincus T, Wolfe F. An infrastructure of patient questionnaires at each rheumatology visit: improving efficiency and documenting care. *J Rheumatol* 2000;27(12):2727-30.
9. Long AF, Scott DL. Measuring health status and outcomes in rheumatoid arthritis within routine clinical practice. *Br J Rheumatol* 1994;33:682-5.
10. Stucki G, Sangha O. Clinical quality management: Putting the pieces together. *Arthritis Care Res* 1996;9(5):405-12.
11. Wolfe F, Pincus T. Listening to the patient: a practical guide to self-report questionnaires in clinical care. *Arthritis Rheum* 1999;42:1797-808.
12. Bellamy N, Kaloni S, Pope J, Coulter K, Campbell J. Quantitative rheumatology: a survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. *J Rheumatol* 1998;25:852-8.
13. Bellamy N, Muirden KD, Brooks PM, Barraclough D, Tellus MM, Campbell J. A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. *J Rheumatol* 1999;26:1593-9.
14. Pincus T. Documenting quality management in rheumatic disease: are patient questionnaires the best (and only) method? *Arthritis Care Res* 1996;9:339-48.
15. Prevoo MLL, Van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44-8.
16. Van Gestel AM, Prevoo ML, Van 't Hof MA, Van Rijswijk MH, Van de Putte LB, Van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum* 1996;39:34-40.

17. Van Riel PL, Schumacher HR. How does one assess early rheumatoid arthritis in daily clinical practice? *Best Pract Res Clin Rheumatol* 2001;15:67-76.
18. Wolfe FW, O'Dell JR, Kavanaugh A, Wilske K, Pincus T. Evaluating severity and status in rheumatoid arthritis. *J Rheumatol* 2001;28:1453-62.
19. Welsing PMJ, Swinkels, HL, Van Gestel AM, Van Riel PLCM. The relation between disease activity and radiologic progression of joint damage in rheumatoid arthritis. *Arthritis Rheum* 2001;44(9 suppl) Abstract 1349.
20. Van Ede AE, Laan RF, Rood MJ, Huizinga TW, Van de Laar, MA, Van Denderen CJ et al. Effect of folic or folinic acid supplementation on the toxicity and efficacy of methotrexate in rheumatoid arthritis: a forty-eight week, multicenter, randomized, double-blind, placebo-controlled study. *Arthritis Rheum* 2001;44:1515-24.
21. Den Broeder AA, Creemers, MCW, Van Gestel AM, Van Riel PLCM. Dose titration using the Disease Activity Score (DAS28) in rheumatoid arthritis patients treated with anti-TNF α . *Rheumatology (Oxford)* 2002;41(6):638-42.
22. Stucki G, Liang MH, Stucki S, Brühlmann, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. *Arthritis Rheum* 1995;38:795-8.
23. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis.* 1987;40:171-8.
24. Hauselmann HJ, Caravatti M, Seifert B, Wang K, Bruckner P, Stucki G, et al. Can collagen type II sustain a methotrexate-induced therapeutic effect in patients with long-standing rheumatoid arthritis? A double-blind, randomized trial. *Br J Rheumatol* 1998;37:1110-7.
25. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol* 1994;140(3):279-89.

Chapter 2

The merits of monitoring: Should we follow all our RA patients in daily practice?

Jaap Fransen¹, Gerold Stucki² and Piet van Riel³.

¹ Department of Rheumatology and Institute of Physical Medicine,
University Hospital Zurich, Switzerland

² Department of Physical Medicine and Rehabilitation,
University Hospital Munich, Germany

³ Department of Rheumatology, University Medical Centre Nijmegen,
The Netherlands

Rheumatoid Arthritis (RA) is mainly characterised by symmetric erosive synovitis and occasional multisystem involvement. Most patients exhibit a chronic fluctuating course of disease that, if left untreated, results in progressive joint destruction, deformity, disability and premature death. It frequently affects patients in their most productive years, and thus, disability results in a major economic loss [1]. There is general agreement that rheumatoid inflammation should be controlled as soon as possible, as completely as possible, and that control should be maintained for as long as possible, consistent with patient safety [2]. It is recommended [3] that, in nearly all cases, patients should be treated with a DMARD or “biologic” agent. DMARDs should be used in high doses, capable to reduce inflammation, unless full treatment effect is gained at lower dosage or limiting toxicity is reached. Treatment failure then follows a simple definition: “treatment failure exists when rheumatoid inflammation is not controlled” [3]. When adequate control is not achieved, the DMARD should be changed or another DMARD or “biologic” agent added [3]. Unfortunately, most patients achieve only partial suppression of rheumatoid inflammation and many lose therapeutic benefit after an initial good response. Additive combination therapy may also produce only temporary benefit [2]. In general, we cannot predict with enough certainty the disease course and occurrence of response and toxicity for individual RA patients. While enriched by new treatment options, the management of RA continues to be a challenge [2,4].

Accepting that the goal of treatment is to reach optimal control of rheumatoid inflammation or even remission, it is clear that management of RA should include systematic and regular evaluation of rheumatoid inflammation. Treatment efficacy should be monitored with the same seriousness as the monitoring of toxicity [5-7]. The treatment program can be accommodated if necessary, from both perspectives of benefit and harm [1]. Such a trade-off could be a good moment to include patient preferences and to educate the patient about treatment options. In daily clinical practice, regular and systematic monitoring of inflammatory activity is useful to:

- Understand if the therapy chosen is needed and effective.
- Assure that rheumatoid inflammation is still under control.
- Make sure that no over treatment is performed.
- Identify rapidly advancing disease with high levels of inflammatory activity over time and fast radiographic progression, where “aggressive” treatment may be needed [3].

- Support the choice of specific DMARDs. It is advised to take the most effective DMARD first, but there is no simple rule as to the order of treatments [3].
- Adjust DMARD dosage in the titration of disease activity.
- Support treatment expectations. In some instances, a full response may take longer than expected, and it may be appropriate to continue the therapy if an adequate response may be achieved by additional treatment time [3].

An example of monitoring for the titration of inflammatory activity is depicted in the graph. The usefulness and feasibility of dose titration in RA patients treated with anti-TNF α is shown in a study presented by Den Broeder et al. [8] in this issue of Rheumatology, demonstrating the advantages of tailoring anti-TNF α treatment compared to the “one size fits all” dosing scheme. Initiatives to establish computerised decision support systems for the monitoring of RA have emerged in Switzerland and Sweden [4,9,10]. Generally, monitoring and documentation is not used regularly in daily clinical practice [11-13]. The statement by Pincus may be recognised: 'Many, if not most, clinicians feel it is not necessary to document obvious clinical improvement recognised by both the patient and the clinician. A positive response to the question “How are you doing?” is sufficient' [14]. In daily clinical practice, there are several barriers to regular monitoring that may not be different from the barriers for the use of outcome measures. Examples are: the apprehension of non-laboratory data as “soft”, unfamiliarity with scores and difficulty with their interpretation, uncertainty about the impact on clinical care or health, the resource of time and personnel, and fear to annoy patients with measurements [15-18]. It is important to recognise these barriers, but they are manageable. By far the most important barrier is: how does a clinician know when rheumatoid inflammation is optimally controlled or in remission? which measures should be used? what are the criteria to state that inflammatory activity is under control?

There is general agreement regarding the measures and examinations that are most appropriate to evaluate change in randomised controlled trials evaluating DMARDs [19-22]. In daily clinical practice however, the aim is to determine and evaluate actual status, rather than a change in status [19, 23]. There are no defined gold standards of severity for the measures used for assessment of RA. Thus, it is not clear “how low you have to go” in which measures, to be quite confident that there is remission or that inflammatory activity is under control. However, the measures that clinicians would apply in daily clinical practice preferably include the American College of Rheumatology (ACR) core

set measures, for practical and methodological reasons. Even if those measures are approximating rather than measuring the underlying synovitis. The two approaches most established to measure change in randomised controlled trials (RCTs) are the ACR improvement criteria and the EULAR response criteria [21,22].

The ACR improvement criteria (e.g. ACR20) determine a patient a responder if there is improvement (e.g. >20%) in both tender and swollen joint counts, and in 3 of the following 5 measures: pain, patient global assessment, physician global assessment, disability, and an acute phase reactant [21]. The ACR improvement criteria are designed to discriminate placebo from verum in RCTs, they are not helpful in assessing actual status in clinic patients [19]. The main disadvantage of such a response measure is that the amount of inflammatory activity you end up with is unknown. To give further insight into status and prognosis, percentile methods for core set measures have been developed, with which severity status of an individual patient can be compared with a reference group [19]. However, it is a disadvantage when changes in multiple measures must be interpreted at the same time, while it is still difficult to interpret their meaning.

The EULAR response criteria determine the patient a good, moderate or non-responder, dependent on both the magnitude of improvement and the absolute level of the DAS28 reached [22]. The DAS28 is an index that includes the results of swollen joint count, tender joint count, ESR and a general health question. The DAS28 ranges virtually from 0-10, a score below 3.2 is defined as "low disease activity". A low level of DAS28 over time reduces the probability of progression of radiological visible joint damage [24,25]. Thus, the DAS28 is suited to aid in determining and evaluating actual status and change in status, particularly when applied to individual patients with RA. As an example, the DAS28 was used to individually increase the dose of methotrexate (MTX), starting with 7.5 mg/week, in a trial on the effects of the addition of folates [26]. How the DAS28 can be used in a step-down regimen is demonstrated by the study of Den Broeder et al. [8]. Clinical variables, as in the DAS28, are just approximates of the underlying synovitis. Clinical assessments may be complemented by information from patient questionnaires such as the Rheumatoid Arthritis Disease Activity Index (RADAI) [27]. However, it is not possible to rely solely on questionnaire results to base treatment decisions on [28]. To be surer about the long-term course of RA in an individual patient, monitoring can include disability and joint damage.

It is important to monitor inflammatory activity on a regular basis, perhaps every visit, every 3 months, or with every change in DMARD and dose. In any case, inflammatory

activity must not only be documented if the patient presents with clearly active inflammation, but also when improvement occurs or treatment effects are expected. Monitoring of long-term effects (disability and joint damage) may take place every 6 months or annually. An important facilitation for clinical use would be, if the information needed to judge inflammatory activity is at hand shortly after the assessment. Then, relevant decisions can be made in presence of the patient.

In primary care, several RCTs are performed on the effectiveness of computerised decision support systems (CDSS) including measurement-feedback and guidelines, mainly in patients having hypertension, diabetes, or anticoagulation need, but also in patients with arthritis [29-37]. According to these studies, CDSS were generally ineffective in changing physician performance or health outcomes. It is important to realise that monitoring of RA is not an intervention that causes health effects, but medication may. Physician performance is the important link between monitoring and health effects through medication.

Conclusion

The goal of treatment of RA is to control rheumatoid inflammation as soon as possible and as completely as possible, and this control should be maintained for as long as possible. To evaluate if treatment goals are reached and to support accommodations of the treatment program, the management of RA patients in daily clinical practice should include systematic and regular evaluation of rheumatoid inflammation. For determining status and change of individual RA patients in daily clinical practice, core set measures and response criteria may be used with some care.

Acknowledgements

We wish to thank Leanne Pobjoy for her assistance in the preparation of the manuscript.

References

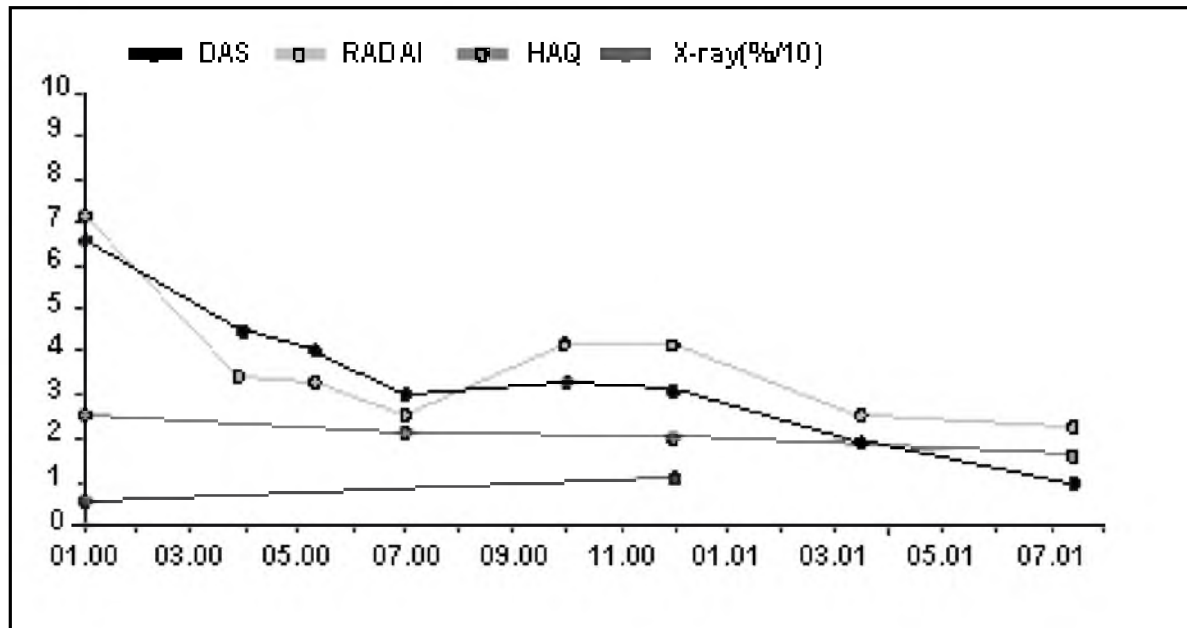
1. Guidelines for the management of rheumatoid arthritis. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. *Arthritis Rheum* 1996;39:713-22.
2. Moreland LW, Russell AS, Paulus HE. Management of rheumatoid arthritis: the historical context. *J Rheumatol* 2001;28:1431-52.
3. Wolfe F, Cush JJ, O'Dell JR, Kavanaugh A, Kremer JM, Lane NE *et al*. Consensus recommendations for the assessment and treatment of rheumatoid arthritis. *J Rheumatol* 2001;28:1413-30.
4. Stucki G, Sangha O. Clinical quality management: putting the pieces together. *Arthritis Care Res* 1996;9:405-12.
5. Wijnands MJ, Van Riel PL. Management of adverse effects of disease-modifying antirheumatic drugs. *Drug Saf* 1995;13:219-27.
6. Simon CH, Vliet-Vlieland TP, Dijkmans BA, Bernelot-Moens HJ, Janssen M, Hazes JM *et al*. Laboratory screening for side effects of disease modifying antirheumatic drugs in daily rheumatological practice. *Scand J Rheumatol* 1998;27:170-9.
7. Guidelines for monitoring drug therapy in rheumatoid arthritis. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. *Arthritis Rheum* 1996;39:723-31.
8. Den Broeder AA, Creemers MCW, Van gestel AM, Van Riel PLCM. Dose titration using the Disease Activity Score (DAS28) in rheumatoid arthritis patients treated with anti-TNF α . *Rheumatology (Oxford)* 2002;41(6):638-42.
9. Uitz E, Fransen J, Langenegger T, Stucki G. Clinical quality management in rheumatoid arthritis: putting theory into practice. *Swiss Clinical Quality Management in Rheumatoid Arthritis. Rheumatology (Oxford)* 2000;39:542-9.
10. Geborek P, Saxne T. Clinical protocol for monitoring of targeted therapies in rheumatoid arthritis. *Rheumatology (Oxford)* 2000;39:1159-61.
11. Wolfe F, Pincus T. Listening to the patient: a practical guide to self-report questionnaires in clinical care. *Arthritis Rheum* 1999;42:1797-808.
12. Bellamy N, Kaloni S, Pope J, Coulter K, Campbell J. Quantitative rheumatology: a survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. *J Rheumatol* 1998;25:852-8.
13. Bellamy N, Muiden KD, Brooks PM, Barraclough D, Tellus MM, Campbell J. A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. *J Rheumatol* 1999;26:1593-9.
14. Pincus T. Documenting quality management in rheumatic disease: are patient questionnaires the best (and only) method? *Arthritis Care Res* 1996;9:339-48.
15. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care* 1989;27(3 suppl):S254-68.
16. Deyo RA, Carter WB. Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint. *Med Care* 1992;30(5 suppl):S176-86.
17. Golden WE. Health status measurement. Implementation strategies. *Med Care* 1992;30(5 suppl):S187-195.

18. Patrick DL: Strategies for improving and expanding the application of health status measures in clinical settings. Discussion. *Med Care* 1992;30(suppl):S198.
19. Wolfe FW, O'Dell JR, Kavanaugh A, Wilske K, Pincus T. Evaluating severity and status in rheumatoid arthritis. *J Rheumatol* 2001;28:1453-62.
20. Boers M, Tugwell P, Felson DT, Van Riel PL, Kirwan JR, Edmonds JP, *et al*. World Health Organisation and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol* 1994(suppl);4186-9.
21. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C *et al*. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
22. Van Gestel AM, Prevoo ML, Van 't Hof MA, Van Rijswijk MH, Van de Putte LB, Van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum* 1996;39:34-40.
23. Van Riel PL, Schumacher HR. How does one assess early rheumatoid arthritis in daily clinical practice? *Baillieres* 2001;15:67-76.
24. Welsing PMJ, Swinkels, HL, Van Gestel AM, Van Riel PLCM. The relation between disease activity and radiologic progression of joint damage in rheumatoid arthritis. *Arthritis Rheum* 2001;44(9 suppl) Abstract 1349.
25. Landewe RBM, Boers M, Verhoeven AC, Boonen A, Van der Heijde DMFM, Van der Linden S. A longitudinal analysis of the relationship between change in DAS28 and change in radiological progression. *Arthritis Rheum* 2001;44(9 suppl) Abstract 737.
26. Van Ede AE, Laan RF, Rood MJ, Huizinga TW, Van de Laar, MA, Van Denderen CJ *et al*. Effect of folic or folinic acid supplementation on the toxicity and efficacy of methotrexate in rheumatoid arthritis: a forty-eight week, multicenter, randomized, double-blind, placebo-controlled study. *Arthritis Rheum* 2001;44:1515-24.
27. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. *Arthritis Rheum*. 1995;38(6):795-8.
28. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. *Rheumatology (Oxford)* 2000;39:321-7.
29. Goldberg HI, Wagner EH, Fihn SD, Martin DP, Horowitz CR, Christensen DB *et al*. A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? *Jt Comm J Qual Improv* 1998;24:130-42.
30. Montgomery AA, Fahey T. A systematic review of the use of computers in the management of hypertension. *J Epidemiol Community Health* 1998;52:520-5.
31. Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. *BMJ*. 2000;320(7236):686-90.

32. Hetlevik I, Holmen J, Kruger O. Implementing clinical guidelines in the treatment of hypertension in general practice. Evaluation of patient outcome related to implementation of a computer-based clinical decision support system. *Scand J Prim Health Care* 1999;17:35-40.
33. Hetlevik I, Holmen J, Kruger O, Kristensen P, Iversen H, Furuseth K. Implementing clinical guidelines in the treatment of diabetes mellitus in general practice. *Int J Technol Assess Health Care* 2000;16:210-27.
34. Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *J Am Med Inform Assoc* 1999;6:104-114.
35. Johnston ME, Langton KB, Haynes B, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med* 1994;120:135-42.
36. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. A systematic review. *JAMA* 1998;280:1339-46.
37. Rubenstein LV, Calkins DR, Young RT, Cleary PD, Fink A, Kosecoff J *et al*. Improving patient function: a randomized trial of functional disability screening. *Ann Intern Med* 1989;111:836-42.

Graph

An example of the monitoring of disease activity (DAS, RADAI), disability (HAQ) and joint damage (X-ray) over time, in an individual RA patient.



Chapter 3

Clinical Quality Management in Rheumatoid Arthritis: ...Putting theory into practice.

Elisabeth Uitz, Jaap Fransen, Thomas Langenegger and Gerold Stucki,
for the members of the Swiss Clinical Quality Management in Rheumatoid Arthritis

Department of Rheumatology and Institute of Physical Medicine,
University Hospital Zurich, Switzerland

© by Rheumatology-Oxford

Abstract

Objective Clinical Quality Management (CQM) in Rheumatoid Arthritis (RA) aims at reduction of inflammatory activity and pain in the short-term, and damage, and consequently disability, in the long-term.

Methods Within CQM as used in Switzerland rheumatologists are provided with a measurement feedback system with which they can regularly follow their patients. Inflammatory activity is measured with the Disease Activity Score (DAS28) and the Rheumatoid Arthritis Disease Activity Index Questionnaire (RADAI), damage with an X-ray score, disability with the Stanford Health Assessment Questionnaire (HAQ). Feedback is used to optimise therapy, which in the short term allows the activity of the inflammatory process to be adjusted or "titrated". In the long term, the therapy result for the individual patient is monitored by the course of disability and damage.

Results In this paper we present a series of cases to illustrate the usefulness of the CQM system in the management of individual RA patients.

Conclusion CQM in RA may be helpful when making decisions about adjustment of treatment, and to document and communicate these decisions based on quantitative data.

Introduction

Quality management focusing on improving health-outcomes may be defined as Clinical Quality Management (CQM) [1]. While effectiveness research examines "what works in medicine", CQM is applying this knowledge to improve care. It aims at a permanent improvement or optimisation of health-outcomes by modifying the process of care in an ongoing learning process [1].

The basis for CQM is the construction of a measurement-improvement system using standardised assessments to be applied in a predefined way. However, there are obstacles to be overcome if CQM methods should become common practice [1]. Practical obstacles, apart from the availability, are: time cost, unfamiliarity with standardised assessments and the meaningful interpretation of measures [2]. The measurement-improvement system should be useful for both clinicians and their patients. Thus, a key issue for successful implementation is that clinicians and patients experience an advantage from measuring standardised clinical and patient-oriented parameters. Most beneficial to the clinician is to have the data at hand when deciding about the therapy of individual patients, and therefore to make immediate use of such data.

The fundamental problem in patients with Rheumatoid Arthritis (RA) is systemic inflammation. The assumptions of the current therapeutic approach are: 1) control of systemic inflammation reduces disease impact to the patient in the short-term, and 2) control of inflammation reduces damage, and consequent disability, in the long-term [3,4,5]. However, now there is increasing evidence that destruction of the joints can progress even in the absence of inflammation. Therefore, efficient control of inflammation may not prevent joint destruction completely [6].

As inflammation is suppressed with potentially toxic drugs, it is critical to find the optimal efficacious dose without introducing intolerable side effects. Especially if a disease-modifying antirheumatic drug (DMARD) has to be stopped and another drug has to be installed, precious time in which the inflammation is not sufficiently suppressed may be lost [7]. Therefore, the expected beneficial effects to the patient have to be monitored with the same conscientiousness as the surveillance for side effects. Nowadays in rheumatology there are measures available for disease impact, damage and disability, and there are response criteria, of which the methodological properties are known [8]. While these measures are now essential when conducting clinical trials [9], they may also be used in daily practice. Using these measures to optimise management in clinical practice leads to a decision on treatment that can be supported by quantitative data, rather than on "personal experience" and overall impression alone.

As a consequence, the activity of the inflammatory process, the primary target of medical therapy, can be adjusted or "titrated", similarly to blood glucose in diabetics or blood pressure in hypertensives.

In this paper we present a series of cases to illustrate the use, and to point out to potential caveats, of a practical CQM system in RA that was introduced to Switzerland in 1997. The CQM allows not only for single patient feedback, but also for provider feedback and health service research.

Methods

Measures

The CQM in RA aims at reduction of inflammatory activity and pain in the short-term and the reduction of damage, and consequent disability, in the long-term. Within this framework, inflammatory activity and damage are intermediate clinical outcomes, while pain and disability are primarily patient-oriented outcomes (see Table 1).

Inflammatory activity was measured with the Disease Activity Score (DAS28) [10,11]. The DAS28 is calculated from the results of a 28 swollen joint count, a 28 tender joint count and ESR (Westergren). The response to treatment can be valued according to the EULAR response criteria, and is dependant on the current level of DAS28 and the magnitude of change [11,12].

Damage was measured with X-rays of the hands and feet, that were scored according to a new method proposed by Rau, the so-called "Ratingen Score", which concentrates on damage to the joint surface [12]. The score was expressed as a percentage from the maximal possible score.

Pain was measured with the self-administered Rheumatoid Arthritis Disease Activity Index (RADAI) questionnaire, which also includes stiffness and global assessment of disease activity [13]. Disability was assessed with the modified Stanford Health Assessment Questionnaire (HAQ) [14,15].

Physicians are provided with manuals about the use of these scores, including information about their validity, reliability, sensitivity to change, and minimal clinical important differences. An extraction of this information is given in Table 2.

Data collection

The physician is provided with a single page record sheet showing two mannequins to mark swollen and tender joints, and space to fill in the ESR and medication information, while patients are provided with questionnaires. The sheets are sent together with the X-ray files to the co-ordination centre. The data are fed into a computer and a feedback report is produced, including graphical displays and tables, which is sent back to the physician. A full example of a feedback report is given in Graph 1, in Graphs 2-4 only the graphical part of the feedback report is presented.

The visits of the patient may be regularly scheduled or be at the initiative of the patient. The measures representing inflammatory activity and pain are measured every visit. Damage and disability are assessed once a year, in a fixed appointment.

Cases

The cases were selected out of the Swiss CQM database to show the use and caveats of the CQM tool. All cases are out-patients, attending on self-admission and have an established diagnosis of RA [16].

Results

Case History 1

In Case 1, it is illustrated how periods of increased inflammatory activity (flares) are presented in graphs and figures, and how the principle of titration of inflammatory activity works in practice.

This 65 year old woman, with a 14 year history of mild RA, treated initially with chloroquine and since 1995 with methotrexate, was followed using standardised assessment since November 1995 (Graph 1).

In 1995, when she volunteered in a trial comparing methotrexate to collagen [17], she developed a flare. With re-installation of methotrexate, disease activity decreased (February 1996). But only with the increase to 10 mg methotrexate (November 1996) was disease activity, as measured with the DAS28 (=1.9) and the RADAI (=0.3), adequately controlled. When methotrexate was stopped in July 1997, one week prior to foot surgery and not re-installed for 6 weeks, the patient had a flare again, with both a high RADAI (=4.9) and DAS (=6.3). With re-installation of methotrexate, disease activity could be titrated at the target of DAS28 (=3.2).

This case illustrates that with DMARD therapy, systemic inflammation can be virtually titrated within several weeks, similar to the titration of blood glucose in a diabetic within hours. It shows that a remission may be best understood as an often temporary state, at the lower end of the continuum of disease activity, which needs to be aimed at continuously through a consequent monitoring and adjustment of treatment. In our case remission with a DAS28 of <2.0 could be achieved once, and a low disease activity with a DAS28 <3.2 was achieved at four time points [10,11]. Cumulated disease activity did not result in damage or disability, as measured with X-rays and the HAQ. Joint destruction is unlikely if the disease activity is low over time [18,19,20].

Keeping patients on continuous adequate treatment is a challenge to the rheumatologist [21]. The graphical display of the clinical status and disease scores is a language that patients can understand and may help to improve compliance. In communication, the use of quantitative data may finally convince our surgical colleagues of the importance to stop DMARDs for just a short period of time with major surgery, or to avoid stopping DMARD treatment at all with minor surgery [22,23].

Case History 2

This case illustrates that while in mild to moderate RA titration of disease activity towards a DAS28 below 3.2 or 2.0 [10] may be achieved using a single DMARD, patients with

more severe RA may require more aggressive treatment, for instance combination therapy.

The 61 year old female patient suffers from RA since 13 years and has been treated with virtually all currently available DMARDs, including parenteral gold, D-penicillamin, methotrexate, sulfasalazine and cyclosporine. Despite these treatments damage progressed, requiring several orthopaedic interventions and leading to a considerable disability, with a HAQ score of 2.3.

With the use of standardised assessments starting in April 1996, it soon became evident that with her current methotrexate therapy alone, in a dose acceptable to the patient (unbearable nausea above the dosage of 17.5mg weekly), disease activity could not be titrated below 3.2 or even 2.0 over a longer period of time (Graph 2). There was an increase in radiological damage from April 1996 to October 1997. An attempted combination therapy with salazopyrine in an increasing dosage over the period of half a year from February to October 1997 was considered not successful. Only the combination with low-dose cyclosporine resulted in an acceptable control of systemic inflammation (DAS28=2.4).

It is illustrated that when faced with figures telling us about a poor prognosis, with an unacceptably ongoing high disease activity and damage, the physician is challenged to consequently and systematically try more aggressive treatment options [24,25]. Trying to attempt acceptable levels of disease activity may be worth while, even in patients with long standing RA and a long history of insufficient treatment.

In terms of measurement this case shows the often congruent course of the semi-objective DAS and the subjective RADAI. In some instances, such as the flare in autumn 1997, the DAS28 but not the RADAI pointed initially to the flare. Having information from both sources at hand may help to clarify in an unclear case, or to ascertain in the case of a congruent course of the indices.

Case History 3

This case, as well as the next case, illustrate some regularly encountered situations requiring individualised interpretation of measures, reminding us that treatment of RA is still an art rather than a cookbook approach.

The 58 year old woman suffers from RA since 7 years. She was first treated with parenteral gold, which had to be stopped because of the development of an allergic rash, then with chloroquine, which was stopped because of inefficacy, and finally, since 1993, with methotrexate.

She was first seen in our clinic in January 1996 when she developed a flare after the cessation of methotrexate treatment. With reinstallation of methotrexate, the initially high DAS28 of 7.3 (Graph 3) receded to just 5.8, which is a response in terms of change (greater than 1.2), but not in terms of the level to be achieved, which should be lower than 3.2 (low disease activity), or even < 2.0 (remission) [10].

As can be seen from Graph 5, where the DAS28 and its components are shown, the number of swollen joints and the ESR decreased, but in the meantime the number of tender joints stayed at a very high level.

A second look at the Graph 3 reveals an extremely high RADAI with values up to 10, as well as a very high HAQ score of 2.4, whereas the X-ray score showed no joint destruction. Remarkably, the HAQ did not change with the values for swollen joints and sedimentation rate.

Indeed this patient did not only suffer from RA, but also from fibromyalgia with a low pain threshold. The discrepancy between the high HAQ as compared to the X-ray score, which was normal, is consistent with a previous report [26]. Patients with fibromyalgia had a higher perceived physical functional disability that was not explained by damage measures. It is also important to note that the DAS includes the number of tender joints, and in this case not only reflects systemic inflammation, but is also driven by a low pain threshold.

It is clear that using standardised assessment and indices requires a cautious look and individualised interpretation [27]. In this patient, disease activity needs to be adjusted based on sedimentation rate and the number of swollen joints, rather than the number of tender joints of the DAS.

Case History 4

In this case we will show that there are situations where one may chose to follow a patient using the course of swollen joints in combination with tender joints and pain.

This 51 year old female patient suffers from RA since 13 years and was successfully treated from 1987 to 1996 with d-penicillamine, and since April 1996 with methotrexate. As the treatment with methotrexate alone did not result in prolonged reduction of disease activity, combination therapy was started.

Firstly with hydroxychloroquine, which was insufficient to control disease activity (flare at 11.1996, see Graph 4). Therefore a triple therapy [25], with additional sulfasalazine was installed in 1997, which induced remission.

It is remarkable in this case that there was no elevation of the sedimentation rate (ESR) above 15 in the 3 flares, but the values for ESR seem to deviate similarly as the DAS, RADAI and pain do. In other words, the ESR remained within the normal range, which may have led to an underestimation of true disease activity by the DAS28 in this patient. A low ESR in active RA could be found in some patients [28].

In Graph 6 there is a congruency seen between swollen joint count and tender joint count, the same congruency is seen in Graph 4 with pain and the DAS28. As in the previous case, the possibility of having an overall picture including information from both the physician and from the patient perspective allows for an ongoing adequate titration of disease activity.

This case illustrates that there is a need to be cautious when using indices such as the DAS28 for individual patients. Most measures are developed based on more or less homogeneous groups of patients. A measure is most likely to be valid for an individual, if that patient represents sufficient characteristics of such a group. But still, for individuals, the disease course and the suitability of a measure to reflect it are difficult to predict.

Discussion

Our cases illustrate the use of a set of standardised assessments within a measurement-improvement system to optimise treatment. As demonstrated, CQM may be helpful when making decisions on the adjustment of treatment, and to document and communicate these decisions based on quantitative data. The tables and graphs used in the feedback report allow the physician a quick overview of status and the course of the disease. The graphs are generally easily understood by the patient, which may enhance compliance with treatment. Positive experiences with the use of the CQM system have been made in Switzerland since 1997. With the participation of all university and large hospital rheumatology units and an increasing number of private rheumatology practices, the CQM concept has gained acceptance.

However, the use of CQM is not trivial and there are some burdens to overcome [2,29]. A problem of standardised assessment for daily practice is the time cost. With the inclusion of patient questionnaires, the effort made is shared by physician and patient [30]. It is important to limit the time costs for the physician to an acceptable level. In our experience, the time costs by the clinical assessments for the physician is regained by the time saved with the availability of useful data. But, the meaningful interpretation of such data requires familiarity with scores, which has to be achieved by regular practice [30]. In addition, reference data from clinical studies may help to interpret clinical status,

or the disease course of a patient. However, the use of such data in clinical practice has not had wide attention. The measures used in the CQM are primarily developed for use in groups, not for individuals. The validity in application for individuals is enhanced by the inclusion of several measures and the inclusion of both clinical and patient-oriented outcomes.

In addition to the demonstrated use of the CQM to adjust drug treatment in individual patients, the cumulated data can be used for provider feedback and group analysis. Rheumatologists are free to use measures at self-defined time-points to adjust treatment, however, they are asked to repeat a complete assessment yearly, including X-rays. The yearly visits serve cohort analysis. The CQM thus allows for the identification of problem areas in treatment, for instance where a high degree of variation in treatment is seen. This may help professional societies to focus on the development of practice guidelines where most benefit can be expected. With a CQM program the implementation of practice guidelines can be monitored. It can be studied whether practice patterns change, and whether, and for whom, this results in hypothesised gains in health-outcomes.

CQM in RA is not a kind of 'cookbook', but a complementary tool to support decision making by the experienced rheumatologist, thus participation has been restricted to rheumatologists only. From a political perspective the concept may thus contribute to further establish the central role of the rheumatologist in the management of RA patients [32,33]. Last but not least, it seems critical that CQM stays within the medical community (here the rheumatology societies) and that the data are confidential. It may be successful only if CQM is not imposed from the top down and if physicians may be sure that their practice will not be scrutinised by third parties. Otherwise it is likely to produce fear, behaviour to protect one's own position, and discrediting of information and its source.

In conclusion, CQM is a potentially useful tool for continuous improvement of patient care. The successful integration into clinical practice is the cornerstone for long-term implementation of CQM. It needs to be studied whether the use of CQM results in improved health outcomes in the short and long term, for patients with RA.

Acknowledgements

We wish to thank Leanne Pobjoy for her assistance in the preparation of the manuscript.

The SCQM project is supported by grants from the Swiss Health Authorities (BAG) and the Swiss Academy for the Medical Sciences (SAMW).

The members of the Swiss Clinical Quality Management in Rheumatoid Arthritis are:

A-M. Chamot, G. Dalvit, A. Forster, P-A. Guerne, P. Hasler, D. Van Linthoudt, BA. Michel, L. Schmid, H. Schwarz, M. Seitz, KL. So, R. Theiler, P. Villiger, P. Wüest.

References

1. Stucki G, Sangha O. Clinical Quality Management: putting the pieces together. *Arthritis Care Res.* 1996;9(5):405-12.
2. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care* 1989;27(3 Suppl):S254-68.
3. Guidelines for the management of rheumatoid arthritis. American College of Rheumatology ad hoc committee on clinical guidelines. *Arthritis Rheum* 1996;39(5):713-22.
4. Fries JF, Williams CA, Morfeld D, Singh G, Sibley J. Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drugs-based treatment strategies. *Arthritis Rheum* 1996;39:616-22.
5. Abu-Skakra M, Toker R, Flusser D *et al.* Clinical and radiographic outcomes of rheumatoid arthritis patients not treated with disease-modifying drugs. *Arthritis Rheum.* 1998;41:1190-5.
6. Bresnihan B. Pathogenesis of joint damage in rheumatoid arthritis. *J Rheumatol.* 1999;26(4):717-9.
7. Stucki G, Langenegger T. Management of rheumatoid arthritis. *Cur Op Rheumatol* 1997;9:229-35.
8. Van Gestel AM, Van Riel PL. Evaluation of early rheumatoid arthritis disease activity and outcome. *Baillière's Clin Rheumatol* 1997;11(1):49-63.
9. Boers M, Tugwell P, Felson DT *et al.* World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol* 1994;41:86-9.
10. Van Gestel AM, Haagsma CJ, Van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis Rheum* 1998;41(10):1845-50.
11. Prevoo MLL, Van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38(1):44-8.
12. Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. *J Rheumatol* 1998;25(11):2094-107.
13. Stucki G, Liang MH, Stucki S, Brühlmann P, Michel BA. A self-administered Rheumatoid Arthritis Disease Activity Index (RADAI) for epidemiological research. Psychometric properties and correlation with parameters of disease activity. *Arthritis Rheum.* 1995;38(6):795-98.
14. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
15. Brühlmann P, Stucki G, Michel BA. Evaluation of a German version of the physical dimensions of the Health Assessment Questionnaire in patients with Rheumatoid Arthritis. *J Rheumatol* 1994;21:1245-9.
16. Arnett FC, Edworthy SM, Bloch DA *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31(3):315-24.

17. Häuselmann HJ, Caravatti M, Seifert B et al. Can collagen type II sustain a methotrexate-induced therapeutic effect in patients with long-standing rheumatoid arthritis? A double-blind, randomized trial. *Br J Rheumatol*. 1998;37(10):1110-7.
18. Van Leeuwen MA, Van Rijswijk MH, Sluiter WJ et al. Individual relationship between progression of radiological damage and the acute phase response in early rheumatoid arthritis. Towards development of a decision support system. *J Rheumatol* 1997;24:20-7.
19. Van Riel PL, Van Gestel AM, Van de Putte LB. Development and validation of response criteria in rheumatoid arthritis: steps towards an international consensus on prognostic markers. *Br J Rheumatol* 1996;35(2 Suppl):4-7.
20. Fries JF, Williams CA, Morfeld D, Singh G, Sibley J. Reduction of long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies. *Arthritis Rheum* 1996;39(4):616-22.
21. Katz JN, Solomon D. Differences between generalists and specialists. (Editorial). *Arthritis Care and Research*, 1997;10(3):161-2.
22. Kasdan ML, June L. Postoperative results of rheumatoid arthritis patients on methotrexate at the time of reconstructive surgery of the hand. *Orthopedics* 1993;16:1233-5.
23. Sany J, Anaya JM, Canovas F, et al. Influence of methotrexate on the frequency of postoperative infectious complications in patients with rheumatoid arthritis. *J Rheumatol* 1993;20:1129-32.
24. Van Gestel AM, Haagsma CJ, Furst DE, Van Riel PLCM. Treatment of early rheumatoid arthritis patients with slow-acting anti-rheumatic drugs (SAARDs). *Baillière's Clin Rheumatol* 1997;11(1):65-82.
25. O'Dell JR, Haire CE, Erikson N, et al. Treatment of rheumatoid arthritis with methotrexate alone, sulfasalazine and hydroxychloroquine, or a combination of all three medications. *N Engl J Med* 1996;334:1287-91.
26. Hawley DJ, Wolfe F. Pain, disability, and pain/disability relationships in seven rheumatic disorders: a study of 1,522 patients. *J Rheumatol* 1991;18:1552-7.
27. Stucki G, Michel BA. How to measure improvement: rules and fallacies. *Rheumatol Europe* 1995;24(2 Suppl):107-11.
28. Wolfe F. Comparative usefulness of C-reactive protein and erythrocyte sedimentation rate in patients with rheumatoid arthritis. *J Rheumatol* 1997;24:1477-85.
29. Deyo RA, Carter WB. Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint. *Med Care* 1992(5 Suppl);30:MS176-86.
30. Wolfe F, Pincus T. Standard self-report questionnaires in routine clinical and research practice -- an opportunity for patients and rheumatologists. (Editorial). *J Rheumatol* 1991;18(5):643-6.
31. Patrick D. Strategies for improving and expanding the application of health status measures in clinical settings. *Med Care* 1992(Suppl.5);30:MS198-201.
32. Stucki G. Specialist management: needs and benefits. *Baillière's Clin Rheumatol* 1997;11(1):97-107.

33. Criswell LA, Such CL, Yelin EH. Differences in the use of second-line agents and prednisone for treatment of rheumatoid arthritis by rheumatologists and non-rheumatologists. *J Rheumatol* 1997;24:2283-90.
34. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Arch Intern Med* 1993;153(11):1337-42.

Table 1 Measurements used in the Swiss Clinical Quality Management in Rheumatoid Arthritis.

	PHYSICIAN	PATIENT (Self-assessment)
Inflammatory Activity	DAS28 - Swollen Joint Count - Tender Joint Count - ESR	RADAI - Pain - Painful Joint Count - Global Assessment - Morning stiffness
Damage and Disability	X-ray Score	HAQ

Table 2 Practical use and interpretation of measurements useful in the titration of disease activity in RA.

Measurement	DAS28	Rx/10	RADAI	HAQ
Practicability				
Time for Physician	5 min.	3 min.		
Time for Patient			+/- 3 min.	+/- 5 min.
Administration*	Short Term	Long Term	Short Term	Long Term
Interpretation				
Variable Type	Continuous	Continuous	Ordinal	Ordinal
Range [†]	0 - 10	0% -100%	0 – 10	0 - 3
Reference Values [‡]	4.1 (1.5)	--	3.0 (1.6, 4.8)	1.0 (0.4, 1.8)
Minimal Clinically Relevant Difference [#]	0.6	3.3%	--	0.17
Target	<= 3.2	No Progression	0	0

* Short term: weeks to months; Long term: 12 months

[†] For all scores: lower scores represent a better health status. For Rx/10, RADAI and HAQ, 0 reflects a normal value. Please note that the lowest border of DAS28 is 0.16, values higher than 9 are seldom seen in practice.

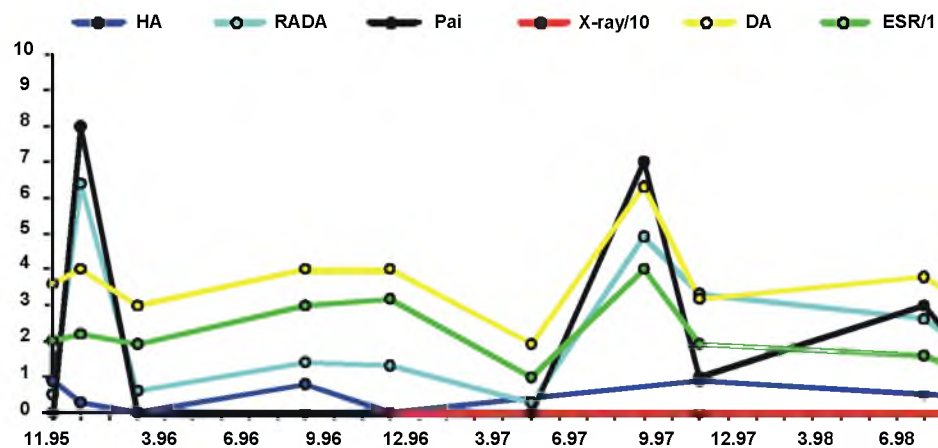
[‡] Mean (sd) or median (25, 75 percentile) in our Swiss CQM population.

[#] References for minimal clinically relevant differences are: [11,13,34].

Feedback Report

SCQM

Physician : Dr. TL
Patient : 1
Date of last visit : 08.98



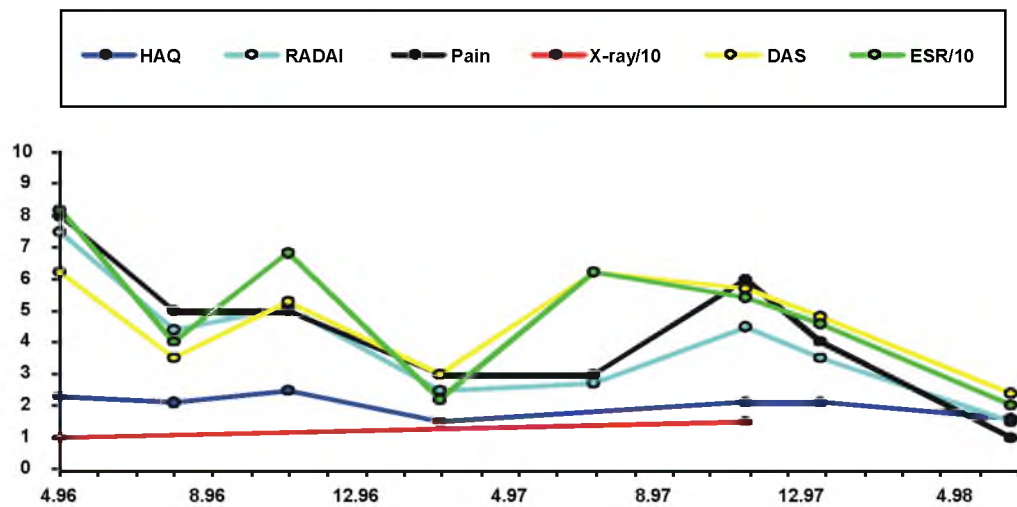
Measure:	11.95	12.95	02.96	08.96	11.96	04.97	08.97	10.97	06.98	08.98
RADA	0.5	6.4	0.6	1.4	1.3	0.3	4.9	3.3	2.6	1.4
Pai	0	8	0	0	0	0	7	1	3	1
DA	3.6	4	3	4	4	1.9	6.3	3.2	3.8	2.8
HA	0.9	0.3	0	0.8	0			0.9	0.5	0.4
X-ray					0			0		

Blood-test:	11.95	12.95	02.96	08.96	11.96	04.97	08.97	10.97	06.98	08.98
ESR (mm/h)	2	2	1	3	3	1	4	1	1	1
Leuco (1000/ul)	3.6	5.24	4	3.6	3.8	3.9	4.5	4.2		
Creatinin (umol/l)	8	9	9	10	10	10	9	10		
Hb (g/dl)	12.	1	11.8	12.	12.	12.4	12.5	11.		
Thrombo (1000/ul)	14	18	16	16	21	21	16	16		
GPT(ALT) (umol/l)	3	2	1	1	1	1	9	3		

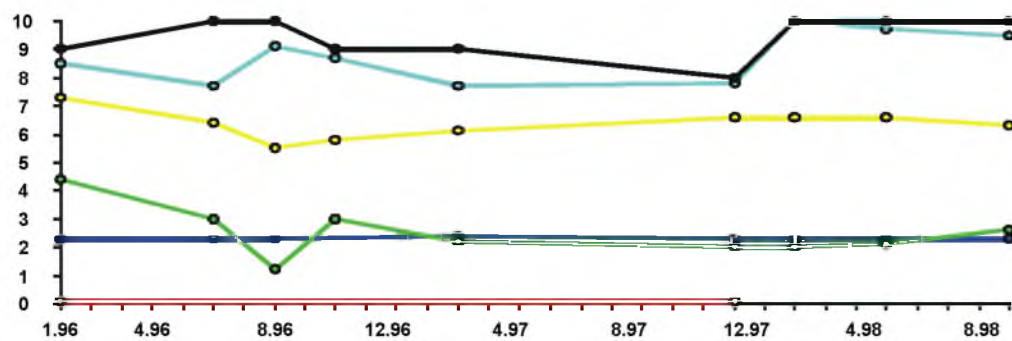
Medication	11.95	12.95	02.96	08.96	11.96	04.97	08.97	10.97	06.98	08.98
Collage	1									
Aulin	10	10	5	20	5		10	10	10	
Methotrexate i.m.		7.5	7.5	7.5	1					
Methotrexate p.o.						1	15	1	17.	17.

* 6 weeks of withdrawal of MTX, reinstallation on 08.97

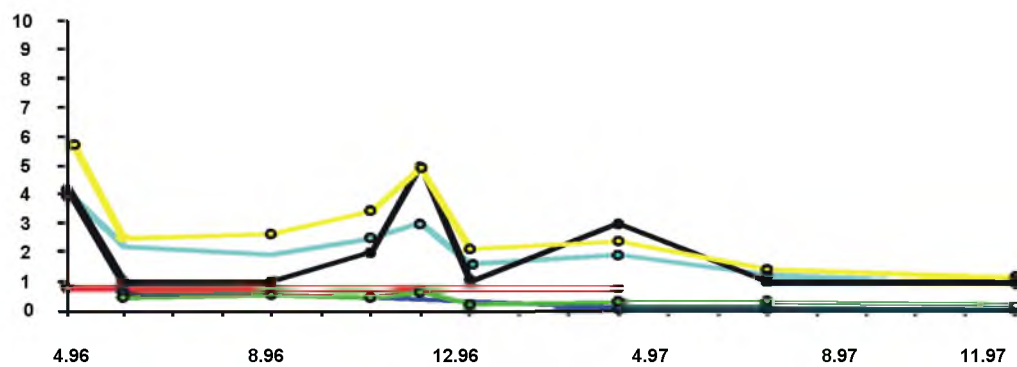
Graph 1. Physician's feedback report of case 1.



Graph 2. Graph from the physician's feedback report of case 2.



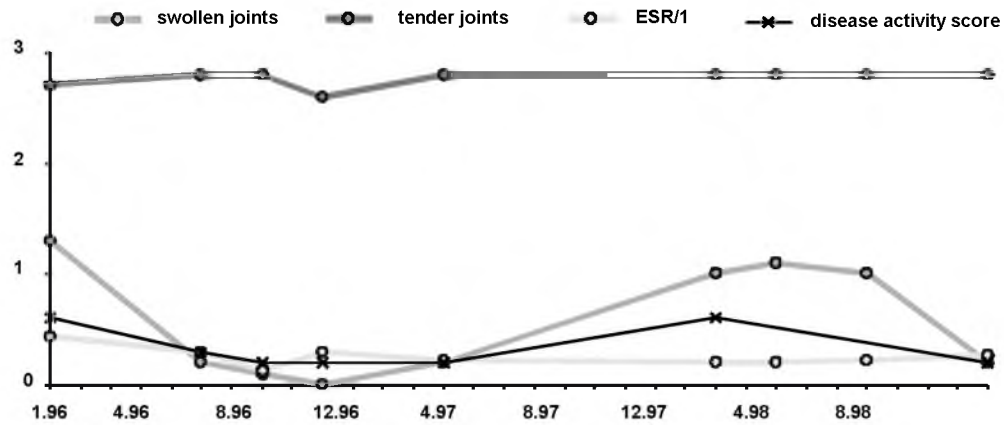
Graph 3. Graph from the physician's feedback report of case 3.



Graph 4. Graph from the physician's feedback report of case 4.

DAS28

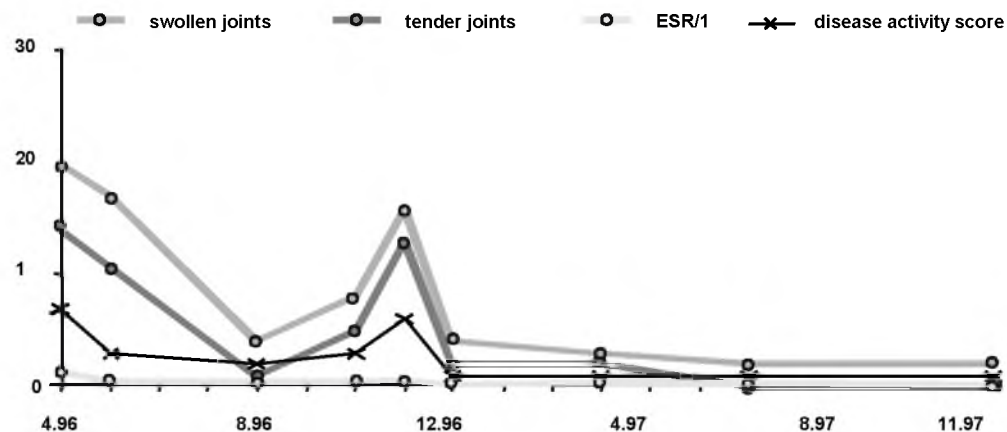
Physician : Dr. TL
Patient : 3
Date of last Visit : 8.98



Graph 5. Graphical representation of the disease activity score (DAS28) and its components for case 3.

DAS28

Physician : Dr. TL
Patient : 4
Date of last Visit : 11.97



Graph 6. Graphical representation of the disease activity score (DAS28) and its components for case 4.

Chapter 4

Feasibility and validity of the RADAI, a self-administered Rheumatoid Arthritis Disease Activity Index.

Jaap Fransen¹, Thomas Langenegger¹, Beat Michel¹ and Gerold Stucki²
for the members of the Swiss Clinical Quality Management in Rheumatoid Arthritis

¹ Department of Rheumatology and Institute of Physical Medicine,
University Hospital Zurich, Switzerland

² Department of Physical Medicine and Rehabilitation, Ludwig-Maximilians University,
Munich, Germany

Abstract

Objective The goal of the Rheumatoid Arthritis Disease Activity Index (RADAI) is to provide an easy to use assessment of disease activity. It is a self-administered questionnaire that combines five items into a single index: current and past global disease activity, pain, morning stiffness and a joint count.

Methods A sample of 584 rheumatoid arthritis (RA) patients was used to assess the internal consistency and the convergent validity of the RADAI.

Results Cronbach's alpha was 0.87, supporting the summation of the items into a single index. The index correlates best with physicians' global assessment ($r=0.59$; $p<0.0001$), HAQ ($r=0.55$; $p<0.0001$) and number of tender joints ($r=0.55$; $p<0.0001$). Correlation with ESR is low ($r=0.27$; $p<0.0001$). The RADAI and DAS28 are correlated ($r=0.53$; $p<0.0001$), but there is low agreement.

Conclusion The RADAI is valid to assess disease activity in RA patients. However, the RADAI may not automatically replace other measures of disease activity, such as the DAS28.

Introduction

The evaluation of the disease course of rheumatoid arthritis (RA) requires comprehensive assessment of process and outcome. As there is no gold standard of disease activity of RA, there are still multiple measures used and needed to assess different aspects of the underlying disease.

For clinical trials in RA there is agreement on which aspects of disease activity have to be assessed at a minimum [1]. According to specific study objectives, measures may be added. This minimal core set contains: an acute phase reactant, a pain rating, a patient's global assessment of disease activity and a physician's global assessment of disease activity, a swollen and a tender joint count, a measure for disability and, for studies lasting longer than one year, a measure for joint damage to the hands and feet [1].

Three measures of the core set, that is, number of swollen joints, number of tender joints and erythrocyte sedimentation rate (ESR), have been integrated in a single index of disease activity, the disease activity score (DAS) [2,3].

Similarly, the items pain and patient global assessment of disease activity are used in a questionnaire format on disease activity, together with other items that are not in the core set such as duration of morning stiffness [4-7]. Whereas joint counts and ESR depend more or less on the physician's judgement, pain and patient global assessment reflect patient perceptions.

The Rheumatoid Arthritis Disease Activity Index (RADAI) [6], is a modification of the questionnaire introduced by Mason [4]. The goal of the RADAI is to provide an easy to use assessment of RA disease activity, which serves as a complement to the physician's assessments and by which the physician's assessment in certain situations could be omitted [6], especially in observational studies or within patient management, where laboratory measurements and clinical assessments may not be possible, or may be too demanding.

The single index approach of the RADAI was found to be valid based on its high association with clinically assessed joint synovitis and the acute-phase response, the high internal consistency, and the loading of the items on a single factor [6]. However, the sample of RA patients was relatively small (N=55).

The objective of the current study is to assess the internal consistency and the convergent validity of the RADAI total score, as well as the feasibility of the questionnaire, cross-sectionally in a larger and broader population than that in which the RADAI was originally tested.

Patients and methods

Patients

We made use of the data from a running cohort of RA patients involved in a clinical quality management project, Swiss Clinical Quality Management in RA (SCQM) [8]. Participating rheumatologists in SCQM come from the university and regional hospitals and private practice. Participating rheumatologists are strongly advised to include all their RA patients, but are free to decide for themselves as to which patients they include. Patients may be included if they have a diagnosis of RA according to the 1987 revised ACR criteria [9] and give their written consent. Prerequisite is the ability to communicate in either French, German or Italian.

Data collection

The inclusion visit includes clinical examination and the taking of a blood sample. Questionnaires are filled in by the patient at home and sent back to the treating physician. The data are collected by the physician and sent to the co-ordination centre. The co-ordination centre provides the physician with a feedback report. The data are stored anonymously to serve for further feedback. Data are only accepted by the centre if all forms are present.

Variables

The patient's personal data include gender, date of birth and, if available, the date of diagnosis of RA according to the patient record. The laboratory measures include rheumatoid factors (RF) (Waalser-Rose or Singer-Plotz) and Westergren erythrocyte sedimentation rate (ESR).

The clinical measures include a 28 swollen joint count, a 28 tender joint count [3] and a global assessment of disease activity by the physician, on an anchored 11 point numerical rating scale (11-NRS). The patient provides a pain rating (11-NRS) and a global assessment of disease activity (11-NRS). The Disease Activity Score (DAS28) was calculated from the results of the 28 swollen joint count, the 28 tender joint count and ESR [3,10]. The questionnaires filled in by the patient included the RADAI [6] and a German version of the Stanford Health Assessment Questionnaire (HAQ) [11,12].

Contents of the RADAI

The RADAI is a 5 item questionnaire. The items ask the patient about 1) global disease activity in the last 6 months, 2) disease activity in terms of current swollen and tender joints, 3) arthritis pain, 4) the duration of morning stiffness and 5) tender joints to be rated in a joint list. The joint list asks about pain in the left and right shoulders, elbows, wrists, fingers, hips, knees, ankles and toes. The first three items are all rated on an anchored numerical rating scale from 0 to 10, where higher scores indicate more disease activity. The scores on the last two items range from 0-6 and 0-48 respectively, but are transformed on the same scale of 0-10. If all items are answered, the scores are added and divided by the number of items to provide a single index of patient-assessed disease activity. (See table 2).

Data processing

The data are stored in an Access 7.0 relational database and have been processed with SAS 6.11 statistical software package.

Statistical analysis

For the description of differences in the population, the two sample T-test or two-sample Wilcoxon test were used for continuous data, the Chi-square test was used when comparing numbers of patients.

Feasibility was assessed descriptively by the number of missing RADAI scores and the number of missing items.

Validity was assessed by a) the correlation of the single RADAI items with core set measures. Expected were moderate correlations and a low correlation for ESR. For all correlations, Spearman's correlation coefficient was used; b) the internal consistency using Cronbach's alpha. An internal consistency of 0.80 is considered adequate on a group level [13]; c) the correlation of the RADAI total score with the scores on the core set measures. It was expected that the highest correlations were to be found with the (semi-objective) tender joint count and the (patient questionnaire) HAQ, and the lowest with ESR; d) the relation of the RADAI to the DAS28, by correlation and by analysing differences of RADAI scores between groups stratified by DAS28-levels [3], ranging from being in remission to high disease activity, with use of Kruskal-Wallis test. The correlation of RADAI with DAS28 was expected to be moderate, as the DAS28 includes ESR. Probability values lower than 0.05 (two-sided) were regarded as significant.

Results

Patients

After an inclusion period of 16 months, the data of 584 patients with an inclusion visit were entered in the database. Virtually all patients were out-patients. Forty-two percent of the patients consulted a rheumatologist at a university clinic, 30% at a regional hospital and 28% in a private practice. Most of the patients were female (72%). Female patients had a larger median disease duration and there are proportionally less women with early RA as men (see Table 1). Disease duration (time since diagnosis) of the total sample ranged from less than one year to 58 years. The HAQ scores were rather uniformly distributed from 0 to 3, median 1.1 (interquartile range 0.4 - 1.8) with a somewhat higher amount of 0 scores and a gradual falling off near the scale's upper end. The mean score of the DAS28 indicates a moderate level of disease activity (see table 5) [3].

Data completeness

The data that were most incomplete were the date of diagnosis, with a 93% completion rate and the RF laboratory test, with a 96% completion rate.

RADAI scores

The distribution of RADAI scores was positively skewed and showed a slight floor effect (see Figure 1). Fourteen percent (n=81) of the patients had a score lower than 1, no patient had a maximal score. The floor effect means that patients with a low RADAI

score may not be able to adequately express a possible improvement in the future [13]. In Table 2 the median scores per RADAI item are given. The scores on the first three items are, relatively to their possible range, higher than on the last two items. All items show similar distributions as in the total score.

Feasibility

A RADAI score was available in 97% of the 584 patients. Eight patients had all RADAI items missing. Another 12 patients had some items missing, for these cases the RADAI was not calculated. The single items had each a completion rate of 98%.

Validity

a) In Table 3, the correlations of the single RADAI items with measures of the core set are shown. The item on morning stiffness correlates least with all core set measures, all the other RADAI items correlate moderately.

The RADAI items had the least correlations with ESR and swollen joint count. The items on pain and patient global assessment, which are in both the RADAI and the core set, are highly correlated.

b) Cronbach's alpha for assessing internal consistency was 0.87 for the RADAI total score. In Table 4 it can be seen that the items "disease activity today" and "arthritis pain today" had the highest correlations with all other items, followed by "disease activity in the last six months" and the "tender joint list". The internal consistency, represented by α , is raised the most by leaving the "morning stiffness" item out.

c) The correlations of the RADAI total score with the scores of the core set measures are represented in Table 5. The mean DAS28 indicates a moderate level of disease activity [3]. The RADAI total score correlates best with the physician's global assessment, tender joint count and the HAQ. The correlation with ESR is low.

The DAS28 correlates well with the physician's global assessment. Two of the DAS28 elements, the swollen and the tender joint counts, correlate relatively well with physician's global assessment of disease activity themselves, $r=0.65$, $p<0.0001$ for both (not shown).

d) In Table 5 it is shown that the RADAI correlates moderately with the DAS28. In the scatter plot of Figure 2, a wide scatter with an impression of curvilinearity can be seen, the latter is primarily caused by the skewed distribution of the RADAI, whereas the DAS28 has a normal distribution [13]. The correlation of the HAQ with the DAS28, $r=0.51$; $p<0.0001$, was similar to that of the RADAI with the DAS28. The two

questionnaires RADAI and HAQ correlated with $r=0.56$; $p<0.0001$. In Table 6 it is shown that the RADAI scores of the groups according to the DAS28 levels [3] of disease activity are significantly different (Kruskal-Wallis, $p<0.0001$).

Discussion

The goal of the RADAI is to provide an easy to use assessment of disease activity in RA patients, especially in observational studies or within patient management, where laboratory measurements and clinical assessments may not be possible or may be too demanding [6]. The results of this cross-sectional study show that the RADAI is feasible in its use by patients and can validly assess disease activity, in a broad sample of RA out-patients.

In terms of feasibility, a vast majority of patients can complete the questionnaire in 5-7 minutes. In this study, the low rate of items that were left blank may indicate that the items were adequate for the patients. In addition, there was a low rate of questionnaires that were not filled in at all, so the acceptance seems to be good.

In terms of validity, the internal consistency of the RADAI is quite high for a questionnaire with only five items [13]. The correlations between the RADAI items and the internal consistency are supporting the summation of the item scores into a total score. The theoretical advantage of the combination of several items into a single index include the higher reliability and reduction of the chance of making a type I error [14]. A problem with an index is that it is less easy to interpret. The RADAI total score has been shown to be able to measure disease activity: it correlates moderately well with core set measures of disease activity where it is expected, and especially relevant, with the DAS28. Further, RADAI scores differ significantly between groups with different disease activity levels according to the DAS28.

With respect to the range of disease duration and the degree of disability based on HAQ scores, this study sample represents a broad selection of RA patients. One advantage of this sample is that it represents RA patients in "real life" circumstances.

The most important study limitation is that the selection of patients by the rheumatologist is not controlled or known, even if rheumatologists were encouraged to include all their RA patients. It is possible that this population represents more severe RA patients. One reason is that general practitioners are not directly involved in the Clinical Quality Management project, but they may see patients with mild RA relatively more often than rheumatologists do.

Most studies on reliability and validity of patient's perception of signs and symptoms of RA are about joint counts. In several studies it has been shown that patient joint counts are reliable [15-17]. The same joint list as used in the RADAI was assessed for test-retest reliability by Stucki [6]. Kappa values ranging from 0.52-0.72 were found for the different joints. Despite adequate reliability within patients, there is no sufficient agreement between patients and clinicians on joint counts [15-17]. For almost the same joint list as used in the RADAI, Hanly [7] found an Intraclass Correlation Coefficient (ICC) of 0.31 for the number of tender joints and ICC of 0.35 for the tender joint score. For that same format, Mason [4] found ICC's ranging from 0.52 to 0.88 for the several joints and an ICC of 0.81 for the total joint score.

The disagreement between patients and clinicians on the rating of joints may be caused by the tendency of patients to rate pain [7] and joint involvement [15-18] higher than clinicians generally do. The different perspective of patients and physicians is illustrated by the findings of Taal [17]: the patient (swollen and tender) joint count correlated with the patient's pain rating, the physician (swollen and tender) joint count correlated with ESR, and both counts correlated with physical disability. Hewlett [18] used separate joint counts for the symptoms pain, heat, swelling and stiffness, and stated that patients clearly can discriminate between them. However, Hanly [7] found that patients cannot rate swelling validly, and no agreement between physician and patient could be found for swollen joint count and swollen joint score.

The reliability and validity of general ratings on pain, global disease activity and morning stiffness is less extensively studied. For the ratings that are identical to the RADAI items on disease activity today, pain and morning stiffness, there was a good test-retest reliability found by Hanly [7], with ICC's ranging from 0.81 to 0.85. The difference with the RADAI format is that a Visual Analogue Scale (VAS) was used. For the Numerical Rating Scale (NRS) as used in the RADAI, Bosi Ferraz [19] found that RA patients can rate their amount of pain as reliably ($r=0.96$) as with a VAS ($r=0.94$).

In concordance with these findings, in the RADAI joint count, the patient is asked to rate pain per joint, but not to rate swelling separate from pain. Current pain, morning stiffness, and disease activity (in terms of swelling and tenderness) are separate items. Pain as a major symptom in RA is addressed by two items. The item on "disease activity in the last 6 months" is held globally, and does not want to discriminate between different symptoms.

The DAS28 and RADAI are reasonably well related ($r=0.53$), but this does not automatically mean that the DAS28 (with a swollen and a tender joint count and ESR) can be replaced by the RADAI.

The first reason is, looking at the wide scatter in the plot of RADAI and DAS28 (Figure 2), that it is not likely that there is sufficient agreement to predict RADAI scores out of DAS scores with a small enough confidence interval [20]. Differences in scaling and distribution are likely to prevent sufficient agreement between laboratory, clinical and patient measures in advance.

The second reason is that DAS and RADAI have different contents. The DAS is parallel to a clinical judgement and the RADAI is a reflection of patient's perception of signs and symptoms. It is in general unlikely that there is enough agreement between process estimators, for instance CRP or ESR, and outcome measures as signs and symptoms. Illustratively, one failed to develop a self-report articular index that correlated with plasma viscosity, the latter used as marker of inflammation [18]. Several studies report no or relatively low correlations of patient questionnaires with laboratory determinants of disease activity such as CRP or ESR [6,21-24].

However, in addition to laboratory and clinical assessments of disease activity, the rating of patient perception of signs and symptoms is useful to assess the burden of disease to the patient. Burden of disease in terms of disability can be assessed with instruments such as the HAQ [11] or AIMS-2 [25]. The RADAI is reflecting burden of disease in terms of impairments. In this study, the HAQ was related to RADAI as well as to the DAS28. Based on factor analysis, Mason [26] pointed out that a measure of signs and symptoms (RADAR) provides complementary information to disability (AIMS-2) rather than duplicative. Further, in patient management as well as in trials, concordance of patient perceptions with physicians' judgement will enhance trustworthiness of interpretations about the underlying disease process.

The major goal of the RADAI is the disease evaluation over time in patient management or in clinical studies. If the RADAI, as a measure of disease activity, is sufficient in comparison to the DAS28 in guiding the physician in the management of individual patients must be clarified in the future. For any longitudinal use, it is needed to assess the smallest detectable difference of the RADAI and its sensitivity to change. Also, decisions on possible item reduction or adaptation should be made on the basis of longitudinal validation studies.

Conclusion

The RADAI has shown to be feasible and valid in the assessment of disease activity in a large cross-sectional sample of RA patients. The RADAI as measure of patient perceived disease activity is related with, but may not automatically replace, other measures of disease activity such as the DAS28.

Acknowledgements

We thank Leanne Pobjoy for her assistance in the preparation of the manuscript.

The Swiss Clinical Quality Management in Rheumatoid Arthritis is supported by grants from the Swiss Academy for Medical Sciences (SAMW) and the Swiss Health Authorities (BAG).

The members of the Swiss Clinical Quality Management in Rheumatoid Arthritis are: A-M. Chamot, G. Dalvit, , A Forster, P-A. Guerne, P. Hasler, D. van Linthoudt, L. Schmid, H. Schwarz, M. Seitz, KL. So, R. Theiler, P. Villiger, P. Wüest.

References

1. Boers M, Tugwell P, Felson DT *et al.* World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol* 1994;21:86-9.
2. Van der Heijde DMFM, Van 't Hof MA, Van Riel PLCM, Van de Putte LBA. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* 1993;20:579-81.
3. Van Gestel AM, Haagsma CJ, Van Riel PLCM. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis Rheum* 1998;41:1845-50.
4. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The rapid assessment of disease activity in rheumatology (RADAR) questionnaire. *Arthritis Rheum* 1992;35:156-62.
5. Stucki G, Stucki S, Brühlmann P, Maus S, Michel BA. Comparison of the validity and reliability of self-reported articular indices. *Br J Rheumatol* 1995;34:760-766.
6. Stucki G, Liang MH, Stucki S, Brühlmann, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. *Arthritis Rheum* 1995;38:795-8.
7. Hanly JG, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. *J Rheumatol* 1996;23:1531-8.
8. Stucki G, Sangha O. Clinical quality management: putting the pieces together. *Arthritis Care Res* 1996;9(5):405-12.
9. MacGregor AJ. Classification criteria for Rheumatoid Arthritis. *Baillières Clin Rheumatol* 1995;9:287-304.
10. Prevoo MLL, Van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44-8.
11. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
12. Brühlmann P, Stucki G, Michel BA. Evaluation of a German version of the physical dimensions of the Health Assessment Questionnaire in patients with rheumatoid arthritis. *J Rheumatol* 1994;21:1245-9.
13. Nunnally JC, Bernstein IH. *Psychometric theory*, 3rd ed. New York: McGraw-Hill 1994.
14. Boers M, Tugwell P. The validity of pooled outcome measures (indices) in rheumatoid arthritis clinical trials. *J Rheumatol* 1993;20:568-74.
15. Prevoo MLL, Kuper IH, Van 't Hof MA, Van Leeuwen MA, Van de Putte LBA, Van Riel PLCM. Validity and reproducibility of self-administered joint counts. A prospective longitudinal followup study in patients with rheumatoid arthritis. *J Rheumatol* 1996;23:841-5.
16. Stewart MW, Palmer DG, Knight RG. A self-report articular index measure of arthritic activity: investigations of reliability, validity, sensitivity. *J Rheumatol* 1990;17:1011-5.
17. Taal E, Abdel-Nasser AM, Rasker JJ, Wiegman O. A self-report Thompson articular index: what does it measure? *Clin Rheumatol* 1998;17:125-9.

18. Hewlett SE, Haynes J, Shepstone L, Kirwan JR. Rheumatoid arthritis patients cannot accurately report signs of inflammatory activity. *Br J Rheumatol* 1995;34:547-53.
19. Ferraz M, Queresma MR, Aquino LRL, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with Rheumatoid Arthritis. *J Rheumatol* 1990;17:1022-4.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986 Feb 8;1(8476):307-10.
21. Riemsma RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with Rheumatoid Arthritis. *Br J Rheumatol* 1996;35:755-60.
22. Pouchot J, Guillemin F, Coste J, Brégeon C, Sany J. Validity, reliability, and sensitivity to change of a French version of the Arthritis Impact Measurement Scales 2 (AIMS2) in patients with rheumatoid arthritis treated with Methotrexate. *J Rheumatol* 1996;23:52-60.
23. Houssien DA, McKenna SP, Scott DL. The Nottingham Health Profile as a measure of disease activity and outcome in rheumatoid arthritis. *Br J Rheumatol* 1997;36:69-73.
24. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997;36:551-9.
25. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum* 1992;35:1-10.
26. Mason JH, Meenan RF, Anderson JJ. Do self-reported arthritis symptom (RADAR) and health status (AIMS2) data provide duplicative or complementary information? *Arthritis Care Res* 1992;5:163-72.

Table 1 Population characteristics.

	Males	Females	Total
N	164	420	584
Age, yrs <i>mean (sd)</i>	58 (12)	59 (13)	59 (12)
Disease Duration, yrs <i>median (IQR)</i>	5 (2-13)	8 (3-17) *	8 (3-15)
Disease Duration < 2 yrs, <i>n</i>	56 (34%)	90 (21%) *	146 (25%)
Rheumatoid Factors +, <i>n</i>	110 (67%)	295 (70%)	405 (69%)

* $p \leq 0.001$, for gender differences; IQR = interquartile range.

Table 2 The RADAI scores in 564 RA patients.

RADAI-item	Possible range	Score
1. In general, how active has your arthritis been over the past 6 months?	0-10	4 (2-6)
2. In terms of joint tenderness and swelling, how active is your arthritis today?	0-10	3 (1-6)
3. How much arthritis pain do you feel today?	0-10	3 (1-5)
4. Were your joints stiff when you woke up today? If yes, how long did this stiffness last? no=0; <30 minutes=1; 30 minutes-1 hour=2; 1-2 hours=3; 2-4 hours=4; >4 hours=5; all day=6	0-6	1 (0-2)
5. Please indicate the amount of pain you are having today in each of the joint areas listed. none=0; mild=1; moderate=2; severe=3 Shoulders, elbows, wrists, fingers, hips, knees, ankles, and toes	0-48	10 (4-18)
Total RADAI score	0-10	3.0 (1.6-4.8)

Score values denote median (interquartile range). For the calculation of the total score the items 4 and 5 are standardised on a scale of 0 to 10.

Table 3 Correlations of RADAI items with several indicators of disease activity.

RADAI-item	ESR	Swollen Joints	Tender Joints	Physician Global Assessment	Pain	Patient Global Assessment	HAQ
1 (arthr. 6 months.)	.23	.35	.45	.50	.67	.71	.46
2 (arthr. today)	.22	.35	.47	.52	.81	--	.45
3 (pain today)	.23	.33	.48	.53	--	.81	.51
4 (morning stiffness)	.17	.25	.36	.36	.49	.47	.37
5 (joint list)	.26	.32	.50	.45	.65	.65	.53

All correlations are significant at a level $p < 0.0001$.
 Identical items are marked with --.

Table 4 At the left: correlations between the RADAI items. At the right: correlations (r) of a RADAI item with all other items, Cronbach's alpha (α) of all other items without the single RADAI item.

RADAI-item	2	3	4	5	r	α
1 (arthritis. 6 months.)	0.70	0.65	0.37	0.51	.67	.85
2 (arthritis. today)		0.81	0.46	0.62	.81	.82
3 (pain today)			0.47	0.67	.81	.81
4 (morning stiffness)				0.48	.51	.89
5 (joint list)					.69	.84

$p < 0.0001$ for all values. Scores for r and α are standardised to have unit variance.

Table 5 Scores of several indicators of disease activity and correlations with RADAI and DAS28 total scores.

Measure	Possible Range	Score	RADAI	DAS28
ESR		20 (10-33)	.27	--
Swollen joint count	0-28	4 (1-9)	.39	--
Tender joint count	0-28	3 (1-8)	.55	--
Pain rating	0-10	3 (1-5)	--	.46
Patient global assessment	0-10	3 (1-6)	--	.46
Physicians global assessment	0-10	2 (1-4)	.59	.72
HAQ	0-3	1 (0.4-1.8)	.56	.51
DAS28	0-10	4.3 (1.4)	.53	

Score values are median (interquartile range), except for DAS28 where mean (sd) is given. Note that, as the swollen and tender joint count and ESR are used for calculating the DAS28, and pain rating and patient global assessment for calculating the RADAI, these correlation values were not calculated. All correlations are significant at $p < 0.0001$.

Table 6 DAS28 groups are formed according to the degree of disease activity.

DAS28 group	n	DAS28 mean (sd)	RADAI median (IQR)
DAS28 < 2.0	42	1.3 (0.5)	1.0 (0.5-1.9)
2.0 < DAS28 ≤ 3.2	120	2.7 (0.3)	2.2 (1.0-3.6)
3.2 < DAS28 ≤ 5.1	260	4.2 (0.5)	2.9 (1.8-4.5)
DAS28 > 5.1	136	6.1 (0.7)	4.9 (3.5-6.5)

IQR=interquartile range.

Figure 1 Scoring on the RADAI in this population is skewed and shows a slight floor effect.

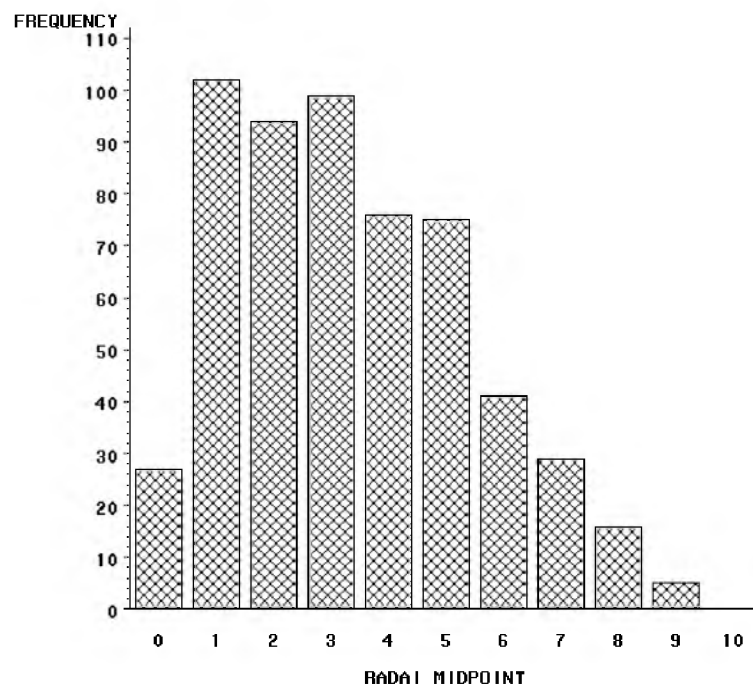
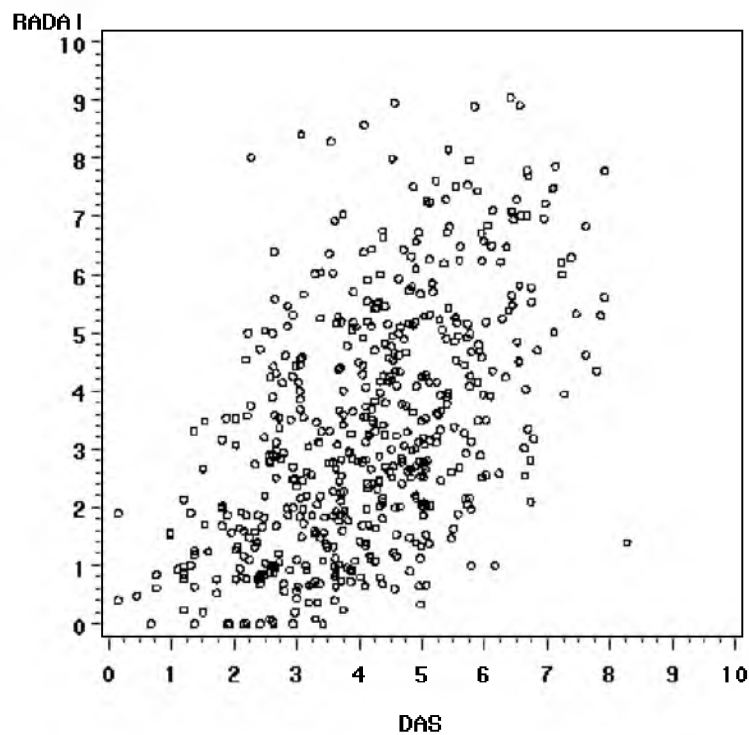


Figure 2 Plot of RADAI against DAS28.



Each observation is marked with a °, observations that are covering each other are indicated by •.

Chapter 5

Responsiveness of the self-assessed Rheumatoid Arthritis Disease Activity Index (RADAI) to a flare of disease activity.

Jaap Fransen¹, Hansjörg Häuselmann², Beat Michel¹, Miguel Caravatti¹, Gerold Stucki³

¹ Department of Rheumatology and Institute of Physical Medicine,
University Hospital Zurich, Switzerland

² Center for Rheumatology and Bone Disease, Klinik im Park, Zurich, Switzerland

³ Department of Physical Medicine and Rehabilitation, University of Munich, Germany

Abstract

Objective The objective of this study is to assess the responsiveness of the Rheumatoid Arthritis Disease Activity Index (RADAI) for increases in disease activity in rheumatoid arthritis (RA) patients, with the occurrence of a flare of disease activity as an external standard.

Methods Post-hoc analysis of data from a randomized double blind controlled trial of MTX versus Collagen II (N=92). Responsiveness is analyzed by the correlation of change in the RADAI with change in the Disease Activity Score (DAS28), and the ability of the RADAI to detect a flare by plotting a receiver operating characteristic (ROC) curve and standardized effect size (SES). The contribution of the single RADAI items to the total change is analyzed by absolute values, the standardized response mean (SRM), and correlation of item score change with the total RADAI score change.

Results The changes in the RADAI correlated highly ($r=0.70$, $p<0.0001$) with changes in the DAS28. The area under the ROC curve was 0.87 (95%CI: 0.78-0.95) for the RADAI, which was similar to the DAS28. The SES for the RADAI was 1.56, which was again similar to the DAS28. The RADAI items on “past global disease activity” and “morning stiffness” contributed the least to the total score change.

Conclusion This study provides evidence that the RADAI is sensitive to detect relevant increases in disease activity in patients with RA. The RADAI may complement clinical measures in clinical studies, or may be used as a proxy for disease activity in epidemiological studies.

Introduction

As there is no ‘gold standard’ of disease activity in rheumatoid arthritis (RA), multiple measures are used and needed in clinical studies [1], as well as for monitoring individual patients in clinical practice for disease management [2]. Progress has been made in rheumatology by the adoption of a ‘core set’ of measures to be used in the trials of antirheumatic drugs [3], and by the use of index measures [4]. With ‘core set’ and index measures, the comparability between studies is enhanced.

The main advantages of indices over single measures are the avoidance of duplicity and increased sensitivity to change [4]. The avoidance of duplicity reduces the risk of making a type I error when performing multiple statistical testing and omits the difficulties arising when trying to interpret different changes in multiple measures simultaneously. Increased sensitivity to change is a result of reduction in measurement error [5,6]. Combination of items to increase validity and reliability is common practice for questionnaires. The main

disadvantages of indices are concerns over validity and practical problems such as interpretation and computational difficulties [4]. The validity of an index depends on the validity of the measures that are included and their appropriate weighting. Interpretation of an index becomes easier when more information from (e.g. discriminative or predictive) validity studies is available and when familiarity with an index is growing.

The Disease Activity Score (DAS), combines erythrocyte sedimentation rate (ESR), a swollen joint count, a tender joint count (and sometimes a general health assessment) into a single index [7]. An index that is self-administered is the Rheumatoid Arthritis Disease Activity Index (RADAI), a short questionnaire that combines five items into a single index: current and past global disease activity, pain, morning stiffness and a joint count. [8,9] The RADAI was developed to complement clinical measures in clinical studies and to serve as a proxy for disease activity in epidemiological studies [9].

While the RADAI has been shown to be a valid measure of disease activity cross-sectionally [9,10], its responsiveness has not been studied. However, a major objective of the RADAI is the measurement of change. Measures that are cross-sectionally valid are not necessarily responsive [11]. Responsiveness is the ability of a measure to capture clinically important changes [11-13].

Until now, most attention concerning responsiveness in RA is implicitly given to changes in direction of improvement, however, in a fluctuating chronic disease such as RA, patients show worsening as well [14]. A measure that showed adequate responsiveness in the direction of improvement, must not necessarily show the same responsiveness in the opposite direction. For constructed measurement instruments as the RADAI, this may depend on the scaling properties, i.e. unidimensionality of the scale [15], item spacing [16], and relative scale length (ceiling or floor effects) [17].

The objective of this study is to assess the responsiveness, or sensitivity to change, of the RADAI for increases in disease activity in patients with RA. As an external standard of change, the occurrence of a flare of disease activity was used. To assess concurrent validity, the RADAI is compared with the DAS28. In addition, to assess content validity, the relative contribution of the individual RADAI items to the amount of RADAI change is assessed.

Patients and methods

Patients

We made use of the data from a randomized placebo controlled trial of methotrexate (MTX) versus Collagen II, that examined whether Collagen II could sustain a MTX induced treatment effect. The study sample and design was described in detail by Hauselmann et al. [18]. Informed consent was obtained from all patients before inclusion. Included were patients with a diagnosis of RA, according to the American College of Rheumatology (ACR) criteria of 1987 [19], and a functional status of class I-III [20]. The patients had to be on unchanged MTX monotherapy for at least 8 weeks. Existing additional medical treatment with glucocorticosteroids or NSAIDs was allowed, if the dose was unchanged for the last 2 weeks and the dose of glucocorticosteroids was ≤ 12.5 mg/day. Excluded were patients younger than 18 years, and patients who had had intra-articular glucocorticosteroid injections within 3 weeks before study begin.

Patients (N=92) were randomized to receive Collagen II with MTX placebo or MTX with Collagen II placebo for 3 months, in a blinded manner. If there was existing additional treatment with NSAIDs or glucocorticosteroids, the doses were kept constant for the entire trial period. The study protocol was approved by the ethical committee of the University Hospital Zurich.

Measurements

The patients were assessed by the same clinician, who was blinded for group assignment, at baseline (t=0), at 30 days (t=1) and at 90 days (t=2). At baseline, sociodemographic data were obtained, radiographs of hands and feet were made and scored according to the Larsen method [21,22]. Laboratory measures included tests for Rheumatic factors (RF), Anti Nuclear Antibodies (ANA), and Human Leucocyte Antigen (HLA) class II typing [23,24], with subsequent classification according to RA association [25]. The following measures were obtained at all visits: Westergren erythrocyte sedimentation rate (ESR) and levels of C-reactive protein (CRP), number of swollen joints (28 joint count), number of tender joints (28 joint count), physician global assessment of disease activity (11-Numerical Rating Scale), pain assessed by the patient (11-NRS), and patient global assessment of disease activity (11-NRS).

The DAS28 (DAS28 S+T) was calculated from the results of the joint counts and the ESR [26]. A German version of the Stanford Health Assessment Questionnaire (HAQ) [27,28] and a German version of the RADAI [9] were used.

The RADAI is a 5 item questionnaire. The items ask the patient about 1) global disease activity during the last 6 months, 2) disease activity in terms of swollen and tender joints today, 3) arthritis pain today, 4) the duration of morning stiffness, and 5) tender joints to be rated in a joint list (see Table 3). The first 3 items are all rated on an anchored numerical rating scale from 0 to 10, where higher scores indicate more disease activity. The scores on the last two items range from 0-6 and 0-48 respectively, but are transformed onto the same scale of 0-10. Demand is that all items are answered. The scores are then added and divided by the number of items to provide a single index of patient assessed disease activity.

The ability of the patient to follow the study protocol was recorded either as regular termination or as discontinuation, with subsequent registration of exact study duration and the specific reason for dropping out. Dropping out was immediately followed by the clinical and laboratory assessments as described before.

A drop-out due to a flare of disease activity was defined as a patient perceived increase in disease activity, leading to a wish to quit the study protocol, and/or a need for a change in medication that caused a protocol violation.

Statistical analysis

Patients who discontinued the study were analyzed for drop-out reasons. Patients with reasons for discontinuation other than a flare of disease activity (e.g. adverse effects), were excluded. Thus, two groups of patients could be formed: patients who discontinued (flare group), and patients who stayed in the study (non-flare group). The groups were formed without classification according to study medication, because splitting up the study sample into four groups would induce a significant power loss. To our knowledge, no differential effects between MTX and Collagen II can be expected in the case of a flare on the clinical parameters that are most relevant for this study, i.e. ESR, number of swollen and tender joints, and patient perception of disease activity. For data with a distribution that was approximately normal (according to the histogram, normal probability plot, median and mean, skewness and kurtosis) parametric statistics were applied and means and standard deviations are presented. For data with a skewed distribution, non-parametric statistics were applied and medians and interquartile ranges are presented.

The flare group and the non-flare group were analyzed for comparability at baseline. Between-group analysis was performed with the Chi-square test or Fisher Exact test, the two sample t-test, or the two sample Wilcoxon test, where appropriate.

The changes between baseline of the RADAI, the DAS28 and the other indicators of disease activity, and the moment of study termination, were compared between both groups. For analyzing within-group differences the one-sample t-test was used, for between-group differences the two-sample t-test as well as the Wilcoxon test were used. Linear regression was used to study the dependency of change in disease activity (DAS28) on disease activity at baseline.

Responsiveness was analyzed in three ways 1) the correlation of change in RADAI with change in the DAS28, by Pearson's correlation coefficient, 2) analysis of the ability of the RADAI to detect a flare, by the plotting of a receiver operating characteristic (ROC) curve (29) and calculation of discriminative ability (area under the curve). The ROC curve represents the sensitivity and specificity of a decision (flare or non-flare) for all possible cut-off points of an underlying variable (change in RADAI score), 3) with a responsiveness statistic, the standardized effect size (SES) [29]. The SES is calculated as the difference of the within-group changes of the flare and the non-flare group, divided by the pooled standard deviations of change. The SES values for RADAI and DAS28 are compared with an approximate z-test [30].

We preferred the use of the DAS28 to test for concurrent validity over the various individual indicators of disease activity for the following reasons: 1) the DAS28 has shown to be a valid estimator of disease activity [26] and is responsive to improvements in disease activity [31]; 2) an index combining 3 items may be a better estimator than 3 items separately; 3) the statistical problem of multiple testing is avoided and interpretation facilitated (3); 4) especially important: for the correlational method, there is no need to develop an a priori expectation to the various degrees of correlation that may be expected with the various indicators of disease activity individually.

The contribution of the single RADAI items to the change in total RADAI score of the flare group is analyzed by the item score change in absolute value, the item responsiveness by use of the standardized response mean (SRM) [32], and correlation of item score change with total RADAI score change by Cronbach's alpha. The SRM is calculated as the ratio of mean change to the standard deviation of change. Higher values indicate a better responsiveness. SRMs are most easily interpreted relative to each other, within the same study.

It is unclear whether different responsiveness statistics, such as SRM [32], effect size (ES), which is calculated as the ratio of mean change to the standard deviation at baseline [33] or the responsiveness statistic proposed by Guyatt, which is calculated as the ratio of a clinically important difference to the variability in stable subjects [11], will

lead to comparable results [34] as was suggested [35]. Therefore, we preferred to use the SES and the SRM, because they represent change and the same population in both nominator and denominator.

The level of statistical significance was set at $p=0.05$ (two-sided). Data were stored in a MedLog database and analyzed with SAS 6.11 statistical software package.

Results

Flare and non-flare group

Within the 90 days of study duration, 26 (57%) of the 46 patients that were on Collagen dropped out. Of the 46 patients on MTX, only 7 (15%) dropped out. No reasons for leaving the study protocol, other than a flare of disease activity, were noted. The median time-to-flare was 31 days (range 17-88; see Figure 1). At baseline, the flare group did not differ significantly in population variables from the patients that continued the study successfully, except for disease duration and ANA positivity (see Table 1). The distribution of gender did not appear to be significantly different between the flare and non-flare group. In addition, there are no statistically significant differences on population parameters within gender. Regarding the clinical parameters at baseline (see Table 2), there is a significant difference between flare and non-flare group in the pain rating only. A high level of DAS28 (31) at baseline was not associated with the later development of a flare (Fisher Exact: $p=0.69$).

Changes within and between groups

At study termination, the non-flare group showed no statistical significant changes from baseline in any indicator of disease activity. But all disease activity indicators were significantly increased (one sample T-test; $p<0.0001$) in the flare group. The changes of the flare group are significantly larger than the non-flare group (Table 2). Comparably, the scores of RADAI and the DAS28 were significantly increased in the flare group, from baseline as well as in comparison to the non-flare group. The changes from baseline have distributions that are normal or close to normal. To interpret the changes from baseline in relation to the median values at baseline, median changes as well as mean changes are given. In the flare group, DAS28 at baseline explained 7% of the variance of change in DAS28, F test: $p=0.012$.

The distribution of score changes in RADAI is depicted in Figure 2. There is some overlap in score changes between the flare and the non-flare group. That means that a part of the patients of both groups share the same amount of change in the RADAI.

Responsiveness

At baseline, the RADAI and the DAS28 correlate with $r=0.58$ (Spearman, $p<0.0001$). As can be seen in the scatter plot (Figure 3), overall, the changes in RADAI are highly related with changes in DAS28; $r=0.70$ ($p<0.0001$).

The discriminative ability of the RADAI for diagnosing a relevant change (a flare) is 0.87, (95% Confidence Interval: 0.78 - 0.95). It is also illustrated by the form of the ROC curve, that is extending towards the upper left corner of the graph (Figure 4). The discriminative ability of the DAS28 regarding the occurrence of a flare is similar: 0.88 (95% CI: 0.81-0.95). The SES of the RADAI $(2.61 - -0.27)/1.84=1.56$ was not significantly different ($p=0.95$) from the SES for the DAS28: $(1.19 - -0.07)/0.80=1.57$. The responsiveness statistics for the several indicators of disease activity are shown in Table 2. Please note that “pain” and “patient global assessment” are included in the RADAI and that the DAS28 includes ESR and the joint counts.

Contribution of single items

The SRMs of the single RADAI items were lower than the SRM of the total score (see Table 3). In absolute values, the items two and three contributed most to the overall change in RADAI score. The items one and four have the lowest SRM and the lowest correlation with change in all other items. Further, the Cronbach's alpha of the RADAI score is raised most by leaving one of these items out.

Discussion

The results of this post-hoc analysis of data from a randomized double blind clinical trial of MTX versus Collagen II provides evidence that the RADAI is sensitive to detect relevant increases in disease activity in patients with RA. The occurrence of a flare of disease activity was used as an external standard of change and was associated with raised values in all disease activity indicators used. Besides the DAS28 and the RADAI, these were measures included in the World Health Organization/International League Against Rheumatism (WHO/ILAR) core set of endpoints for RA clinical trials [3]. The responsiveness, of the RADAI is indicated by 1) the changes in the RADAI are well correlated with changes in the DAS28; 2) the discriminative ability of the RADAI and of the DAS28, reflected by the area under the ROC curve, are virtually the same. The area is significantly different from 0.5, where 0.5 indicates that no information is provided [13]; 3) the SES values of the RADAI and the DAS28 are about the same and larger than one. Interpretation guidelines are not available, but the SES values we found are quite high in

comparison with values found elsewhere. Buchbinder et al. [30] reported on SES for clinical measures ranging from 0.70 (physician global assessment) to 0.07 (ESR), Vliet Vlieland et al. [36] reported values of 0.98 (Ritchie Articular Index) towards 0.30 (duration of morning stiffness).

Of the single RADAI items, the first item “global disease activity in the past 6 months” and the fourth item “duration of morning stiffness” appeared to be the least responsive. However, with a SRM of 0.8 or greater they can be considered to be very sensitive measures [32]. However, from the nature of the question, the item about the level of disease activity in the past six months cannot be expected to be very responsive, even if the impression of the level of past disease activity may be influenced by the disease activity currently experienced.

There may be some practical and economical reasons to replace assessments done by the physician with patient questionnaires. However, the fact that the RADAI and the DAS28 perform similarly well does not mean one automatically can replace the DAS28 with the RADAI [10]. The DAS28 and the RADAI both aim to measure the same underlying construct, arthritic inflammatory activity, but from two different ‘perspectives’ [10]. The perspective of the patient is formed by perceptions of symptoms and signs, the perspective of the physician by judgment of joint swelling, joint tenderness, and ESR. One of the reasons that the RADAI performs well towards the DAS28 in this study may be the use of a flare as an external standard of change. The RADAI may be a more direct measure of this unpleasant experience than the DAS28. On the other hand, the DAS28 may reflect the underlying arthritic inflammatory activity more directly [37].

A study limitation is that a flare of disease activity is a large change for RA patients. It is possible that e.g. morning stiffness may not be very responsive to smaller but relevant changes in disease activity. However, the correlation and scatter plot of RADAI and DAS28 changes are indicating that the RADAI may be sensitive for smaller changes as well. In future studies, the sensitivity of the RADAI for small but clinically relevant changes in disease activity has to be assessed together with its smallest detectable difference.

We choose not to stratify our sample into subsamples according to medication, because power would be lost. We could find no strong indices that patients who had a flare on MTX differed at baseline and in response from patients who had a flare on receiving Collagen II (data not shown). It appears that no important bias has been introduced. The

criterion of a flare was used based on the assumption that the patients had stable disease activity at baseline and that improvements were not expected by patients or physicians. Indeed, the flare group showed significant increases in all indicators of disease activity, whereas the non-flare group showed no changes. A high level of DAS28 at baseline was not predictive for the development of a flare later on. Further, the amount of change in DAS28 was largely independent of the baseline value of DAS28 in the flare group. Thus, it is unlikely that patients dropped out due to persistent high disease activity instead of a flare, or that patients with high levels of disease activity at baseline could not “express” their flare in DAS28.

In contrast to our study, most studies done on responsiveness of measures in RA concentrate on treatment effects and an improvement of disease activity is expected. In few studies, the responsiveness of questions on patient perceptions of signs and symptoms, similar to the RADAI, was studied. In a trial for multidisciplinary treatment in RA, the VAS for severity of morning stiffness was more responsive than the VAS for morning stiffness duration [36]. Moreover, the morning stiffness severity measure was the most responsive of all (core set) endpoint measures used. In a RCT comparing low-dose cyclosporine with placebo, morning stiffness (in minutes) and pain (VAS and Likert scale) were among the least responsive measures. But the combination of four arthritis pain related questions (AIMS-2) appeared to be the only measure more responsive than the standard tender joint count [30]. In a trial of several anti-rheumatic drug treatment strategies, the physical discomfort measures, pain and well-being, were the most responsive for treatment effect at 3 months [39]. In two other studies [8,40], time was used as an external standard of change, and some worsening was likely to have occurred, but was not clearly dealt with in the results. A similar questionnaire format as the RADAI, was tested for agreement on change by computing the ICCs between patients and clinicians for changes in scores over a mean 6-month interval. However, the changes of the questionnaire were not compared with external indicators for change in disease activity, making it uncertain how far clinical relevant changes occurred [8]. In an observational 60-week study on 24 RA patients, it was found that physician and patient global assessments, pain scores, and the HAQ were more sensitive to change than other clinical measures, while laboratory measures were generally less sensitive to change [40]. Thus, it appears that the responsiveness of questions about patient perceptions on signs and symptoms may, in the least, equal the responsiveness of clinical measures. The different findings with morning stiffness may illustrate the importance of the design of

the question and the scaling. To further increase the responsiveness of the RADAI, the morning stiffness item could be changed to measure severity instead of duration [36]. The RADAI item on past disease activity could be replaced by a question regarding general well-being in the context of the arthritis [41].

In summary, this study provides evidence that the RADAI, a self-assessed Rheumatoid Arthritis Disease Activity Index, is sensitive to detect relevant increases in disease activity in patients with RA. In future studies the sensitivity of the RADAI has to be assessed for small but relevant clinically relevant changes in increasing and decreasing disease activity. The RADAI may complement clinical measures in clinical studies, or may be used as a proxy for disease activity in epidemiological studies.

Acknowledgments

We are grateful for the critical reading of the manuscript by Anke van Gestel, MSc. Department of Rheumatology, University Hospital Nijmegen, The Netherlands.

We thank Leanne Pobjoy for her assistance in the preparation of the manuscript.

References

1. Kirwan JR. A theoretical framework for process, outcome and prognosis in rheumatoid arthritis. *J Rheumatol* 1992;19:333-6.
2. Stucki G, Langenegger T. Management of rheumatoid arthritis. *Curr Opin Rheumatol* 1997;9:229-35.
3. Boers M, Tugwell P, Felson DT, Van Riel PL, Kirwan JR, Edmonds JP, et al. World health organisation and international league of associations for rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol* 1994;41:86-9.
4. Boers M, Tugwell P. The validity of pooled outcome measures (indices) in rheumatoid arthritis clinical trials. *J Rheumatol* 1993;20:568-74.
5. Roberts RS. Pooled outcome measures in arthritis: the pros and cons. *J Rheumatol* 1993;20:566-7.
6. Goldsmith CH, Smythe HA, Helewa A. Interpretation and power of a pooled index. *J Rheumatol* 1993;20:575-8.
7. Van der Heijde DM, Van 't Hof MA, Van Riel PL, Theunisse LA, Lubberts EW, Van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis*. 1990;49:916-20.
8. Mason JH, Anderson JJ, Meenan RF, Haralson KM, Lewis-Stevens D, Kaine JL. The Rapid assessment of disease activity in rheumatology (RADAR) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. *Arthritis Rheum* 1992;35:156-62.
9. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. *Arthritis Rheum* 1995;38:795-8.
10. Fransen J, Langenegger T, Michel BA, Stucki G. Feasibility and validity of the RADAI, a self-administered rheumatoid arthritis disease activity index. *Rheumatology Oxford* 2000;39:321-7.
11. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis*. 1987;40:171-8.
12. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985;38:27-36.
13. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991;12 Suppl 4:142S-158S.
14. Fortin PR, Stucki G, Katz JN. Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis Rheum*. 1995;38:1027-30.
15. Silverstein B, Kilgore KM, Fisher WP, Harley P, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: I exploring unidimensionality. *Arch Phys Med Rehabil*. 1991;72:631-7.
16. Silverstein B, Fisher WP, Kilgore KM, Harley P, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: II defining interval measures. *Arch Phys Med Rehabil*. 1992;73:507-18.

17. Nunnally JC, Bernstein IH. "Psychometric theory" 3 ed. McGraw-Hill, New York 1994.
18. Hauselmann HJ, Caravatti M, Seifert B, Wang K, Bruckner P, Stucki G, et al. Can collagen type II sustain a methotrexate-induced therapeutic effect in patients with long-standing rheumatoid arthritis? A double-blind, randomized trial. *Br J Rheumatol* 1998;37:1110-7.
19. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
20. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in RA. *Arthritis Rheum* 1992;35:498-502.
21. Larsen A, Dale K, Eek M. Standard radiographs of rheumatoid arthritis. Department of Radiology, Oslo Sanitetsforenings Rheumatism Hospital, 1975.
22. Larsen A. How to apply larsen score in evaluating radiographs of rheumatoid arthritis in longterm studies. *J Rheumatol* 1995;22:1974-5.
23. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* 1992;39:225-35.
24. Blasczyk R, van Lessen A, Schwella N, Huhn D, Salama A. A novel HLA-DR13 allele (DRB1*1314) identified by single-strand conformation polymorphism analysis and confirmed by direct sequencing. *Hum Immunol* 1995;43:309-12.
25. Weyand CM, Hicok KC, Conn DL, Goronzy JJ. The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. *Ann Intern Med* 1992;117:801-6.
26. Prevoo MLL, Van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum*. 1995;38:44-8.
27. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
28. Bruhlmann P, Stucki G, Michel BA. Evaluation of a German version of the physical dimensions of the Health Assessment Questionnaire in patients with rheumatoid arthritis. *J Rheumatol* 1994;21:1245-9.
29. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis* 1986;39:897-906.
30. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality of life measures' responsiveness to treatment in a randomized controlled trial. *Arthritis Rheum* 1995;38:1568-80.
31. Van Gestel AM, Prevoo MLL, Van 't Hof MA, Van Rijswijk MH, Van de Putte LBA, Van Riel PLCM. Development and validation of the european league against rheumatism response criteria for rheumatoid arthritis. *Arthritis Rheum* 1998;41:1845-50.
32. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632-42.

33. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27 Suppl 3:S178-89.
34. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997;50:239-46.
35. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995;48:1369-78.
36. Vliet Vlieland TPM, Zwinderman AH, Breedveld FC, Hazes JMW. Measurement of morning stiffness in rheumatoid arthritis clinical trials. *J Clin Epidemiol* 1997;50:757-63.
37. Van der Heijde DMFM, Van 't Hof MA, Van Riel PCLM, Van Leeuwen MA, Van Rijswijk MH, Van de Putte LBA. Validity of single variables and composite indices for measuring disease activity in rheumatoid arthritis. *Ann Rheum Dis* 1992;51:177-81.
38. Van Gestel AM, Haagsma CJ, Van Riel PLCM. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis Rheum.* 1998;41:1845-50.
39. Van der Heide A, Jacobs JWJ, Schenk Y, Bijlsma JWJ. Endpoint measures in rheumatoid arthritis clinical trials: group summary and individual patient analysis. *J Rheumatol* 1994;21:2195-9.
40. Ward MW. Clinical measures in rheumatoid arthritis: which are most useful in assessing patients? *J Rheumatol* 1994;21:17-27.
41. Hanly JG, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. *J Rheumatol* 1996;23:1531-8.

Table 1 Population characteristics.

	Non-Flare	Flare	P-value	Total
n=	59	33		92
Male	8 (14%)	8 (24%)	0.20	16 (17%)
Female	51 (86%)	25 (76%)		76 (83%)
Age, years	51 (12)	53 (13)	0.37	52 (13)
Disease Duration, years	7 (3-11)	13 (6-16)	0.0014	9 (4-14)
Body Mass Index, kg/m ²	23 (21-25)	23 (21-26)	0.89	23 (21-25)
Use of Steroids, yes	23 (39%)	18 (55%)	0.17	41 (45%)
Use of NSAIDs, yes	49 (83%)	24 (72%)	0.26	73 (79%)
ANA, positive	39 (66%)	30 (91%)	0.011	69 (75%)
RF, positive	49 (83%)	31 (94%)	0.32	80 (88%)
HLA-DRB1 A-D, positive	31 (62%) ^a	19 (66%) ^b	0.35	50 (63%)
X-ray, range 0-5 points	2.3 (2.0-3.1)	2.5 (2.1-3.1)	0.27	2.3 (2.1-3.1)
ACR I (n=)	13 (22%)	3 (9%)	0.26	16 (17%)
Functional Class II	16 (27%)	12 (36%)		28 (31%)
III	30 (51%)	18 (55%)		48 (52%)
IV				0

Population characteristics and between-group differences at baseline. Values are counts (column percentage), mean (SD) for age, median (interquartile range) for the other variables, that are skewed. P-values are concerning between group differences.

ANA = Anti Nuclear Antibodies, RF = Rheumatic factors, HLA-DRB1 A-D = RA associated Human Leucocyte Antigen alleles, classification according to Weyand [23]. Due to missing values: ^a n=50, ^b n=29

Table 2 Changes in measures of disease activity

Measure	Range	Baseline		p	Median change from baseline *		Mean change from baseline *		Responsiveness		
		Non Flare	Flare		Non Flare	Flare	Non Flare	Flare	r	pa	SES
CRP mg/l		12 (5 – 29)	11 (4 – 25)	0.62	-1 (-5 – 6)	+18 (7 – 75)	-1.1 (35)	+ 39 (46)	0.34	0.84	1.03
ESR mm/h		16 (8 – 33)	18 (12 – 25)	0.91	0 (-4 – 6)	+10 (4 – 31)	0 (13)	+19 (20)	0.38	0.81	1.18
Swollen Joints	0-28	5 (2 – 9)	5 (3 – 10)	0.46	0 (-2 – 1)	+3 (0 – 6)	-0.3 (3.0)	+3.4 (3.8)	0.41	0.78	1.14
Tender Joints	0-28	5 (2 – 8)	6 (2 – 9)	0.54	0 (-2 – 1)	+3 (2 – 8)	-0.6 (4.5)	+4.4 (4.8)	0.61	0.80	1.08
Pain	0-10	3 (2 – 5)	4 (3 – 6)	0.032	0 (-1 – 1)	+3 (2 – 5)	-0.2 (2.0)	+3.3 (2.4)	0.79 ^a	0.89	1.67
Patient Global	0-10	3 (2 – 5)	4 (3 – 6)	0.32	0 (-1 – 1)	+3 (2 – 5)	-0.3 (2.3)	+3.1 (2.3)	0.84 ^a	0.87	1.51
Physician Global	0-10	3 (2 – 4)	3 (3 – 5)	0.23	0 (-1 – 1)	+3 (2 – 5)	0 (1.3)	+2.0 (1.6)	0.61	0.84	1.46
HAQ	0-3	1.0 (0.4 – 1.5)	1.3 (0.5 – 1.8)	0.38	0 (-0.4 – 0.1)	+0.4 (0.2 – 0.7)	-0.1 (0.5)	+0.47 (0.5)	0.77	0.80	1.16
RADAI	0-10	2.0 (0.8 – 3.7)	2.8 (1.5 – 4.7)	0.082	-0.3 (-0.9 – 0.6)	+2.6 (2.0)	-0.3 (1.7)	+2.6 (2.0)	1.00	0.88	1.57
DAS28	0-9	4.3 (1.4)	4.3 (1.3)	0.89			-0.1 (0.8)	+1.2 (0.8)	0.70	0.87	1.56

Measures of disease activity for both groups, at baseline and differences from baseline at the moment of study termination. At baseline, values are mean (SD) for the

DAS28 and median (interquartile range) for the other variables. Changes from baseline are given in median (IQR) and in mean (SD). A positive sign (+) indicates an increase, for all measures this indicates higher disease activity. A negative sign (-) indicates a reduction. P-values are concerning between group differences.

Responsiveness statistics: r = correlation of change with change in RADAI, pa = predictive ability for detecting a flare, SES = standardized effect size.

All (Pearson's) r are significantly different ($p < 0.001$) from 0, all pa are significantly different ($p < 0.05$) from 0.5.

^a Item correlation with the RADAI, after deletion of that item from the RADAI

* P-values for between-group differences are ≤ 0.0001 for median changes (Wilcoxon) and ≤ 0.0001 for mean changes (T-test)

Table 3 Contribution to change

RADAI Items	Mean change	SRM	α	r
1. In general, how active has your arthritis been over the past 6 months?	2.5 (3.2)	0.78	0.84	0.43
2. In terms of joint tenderness and swelling, how active is your arthritis today?	2.9 (2.6)	1.12	0.72	0.84
3. How much arthritis pain do you feel today?	3.5 (3.0)	1.17	0.73	0.79
4. Were your joints stiff when you woke up today? If yes, how long did this stiffness last? (no=0; <30 minutes=1; 30 minutes-1 hour=2; 1-2 hours=3; 2-4 hours=4; >4 hours=5; all day=6)	2.2 (2.7)	0.81	0.83	0.47
5. Please indicate the amount of pain you are having today in each of the joint areas listed below. (none=0; mild=1; moderate=2; severe=3; shoulders, elbows, wrists, fingers, hips, knees, ankles, toes.)	1.9 (1.6)	1.19	0.79	0.58
Total RADAI score	2.6 (2.0)	1.31	0.82	--

The change scores of the single RADAI items of the flare group (n=33). Mean (SD) of change, the standardized response means (SRM), Cronbach's alpha (α) of the RADAI without the item, the correlation (r) of change in the item with change in all other items.

For the total RADAI score, the Cronbach's alpha (α) is on all items.

Figure 1 Number of drop-outs due to flare

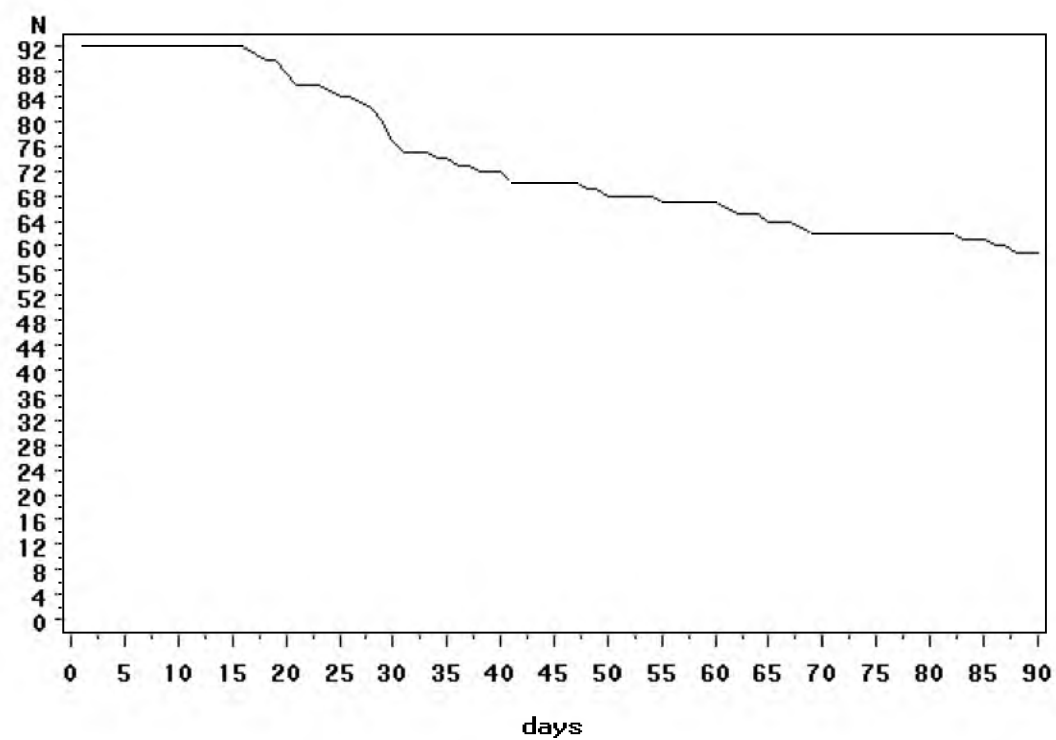


Figure 2 Distribution of the changes in RADAI

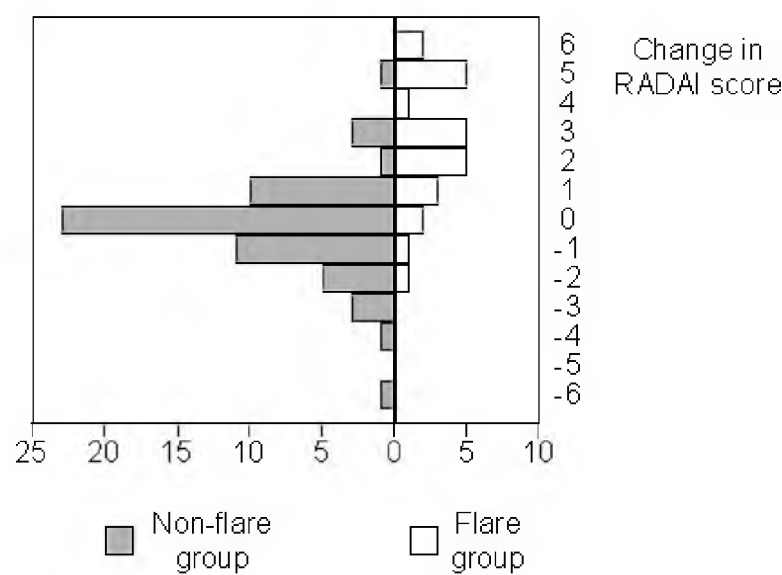


Figure 3 Scatter plot of change in RADAI against change in DAS28

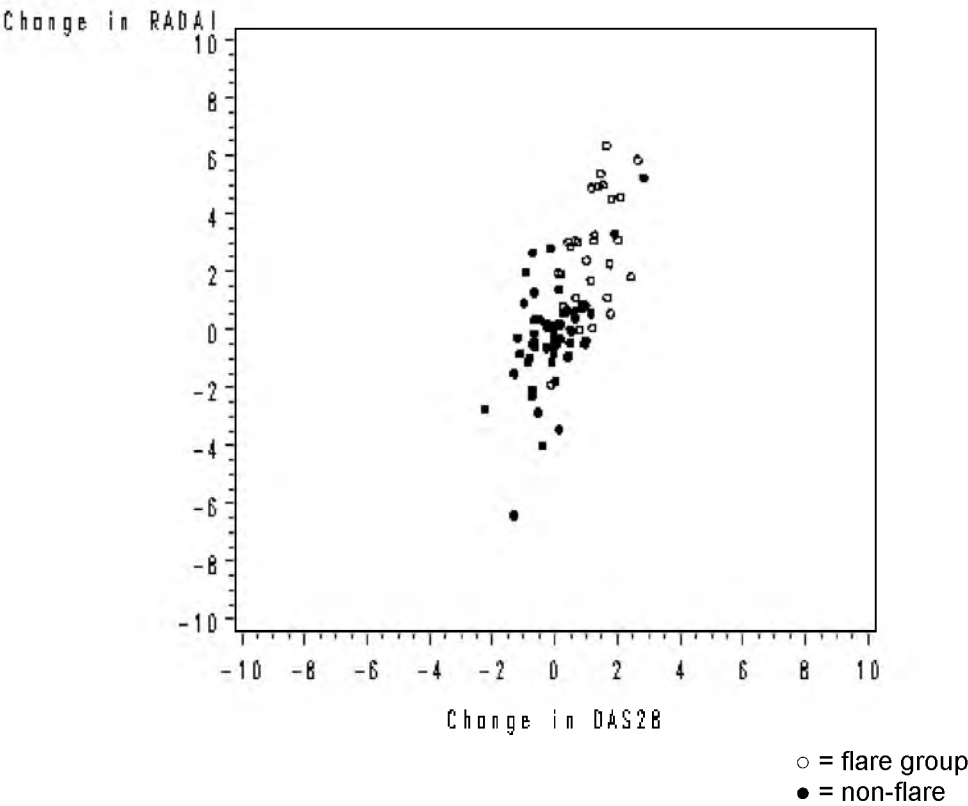
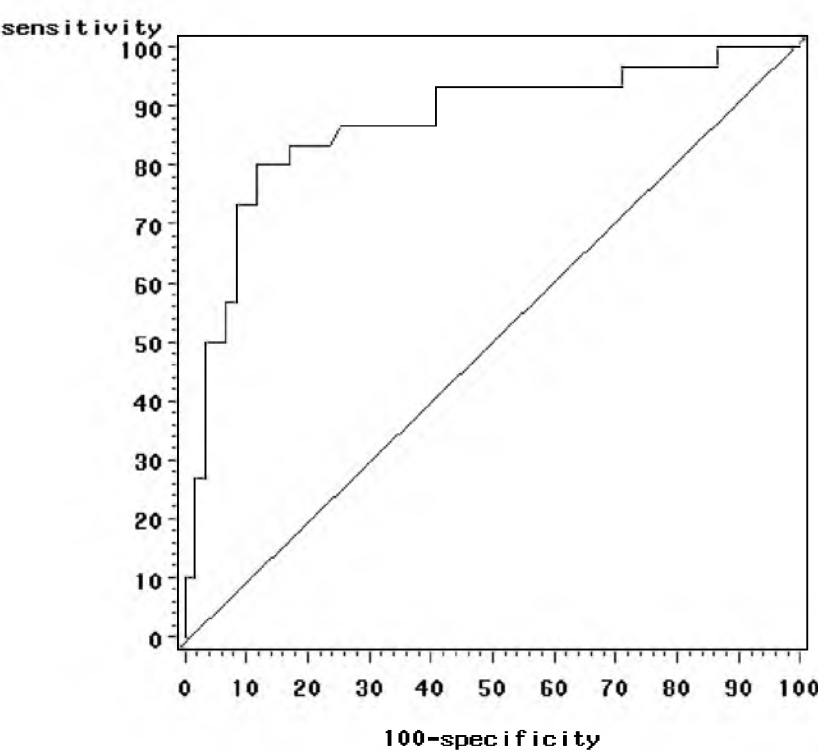


Figure 4 Receiver Operating Characteristic curve of changes in RADAI



Chapter 6

The effectiveness of a measurement-feedback system on outcome in Rheumatoid Arthritis: a controlled clinical trial.

Jaap Fransen¹, Gerold Stucki², Jos Twisk³, Anne-Marie Chamot⁴, Jean-Charles Gerster⁵,
Thomas Langenegger⁶, Michael Seitz⁷, Beat Michel¹
and the members of the Swiss Clinical Quality Management
in Rheumatoid Arthritis (SCQM).

¹ Department of Rheumatology and Institute of Physical Medicine,
University Hospital Zurich, Switzerland

² Department of Physical Medicine and Rehabilitation,
University Hospital Munich, Germany

³ Institute of Research in Extramural Medicine, Faculty of Medicine,
Free University Amsterdam, The Netherlands

⁴ Private practice, Morges, Switzerland

⁵ Department of Rheumatology, Physical Medicine and Rehabilitation, University Hospital
Lausanne, Switzerland

⁶ Adelheid Page Rehabilitation Hospital, Unteraegeri, Switzerland

⁷ Clinic for Rheumatology and Clinical Immunology and Allergology,
University Hospital Berne, Switzerland

Abstract

Objective With the help of a measurement-feedback system, the treatment strategy for individual RA patients can be adjusted to achieve optimal control of disease activity. The main objective was to study whether a measurement-feedback system is effective in reducing disease activity in RA patients.

Methods Forty-eight rheumatologists and 264 patients participated in a controlled clinical trial. A 3-months control period was followed by a 12-months period, where feedback on disease activity, disability and damage was provided to the rheumatologist. Primary outcome measure was the Rheumatoid Arthritis Disease Activity Index (RADAI).

Results For 62% of the patients, the feedback system was used. DMARD changes occurred in 41% of the patients. In patients with high disease activity and feedback use ($n=70$), the RADAI decreased in the feedback period with -0.27 point per 30 days ($p<0.05$), as compared to the control period. Patients for which the feedback system was used had a better outcome than non-users.

Conclusion Much more training regarding the use of a feedback system and outcome measures, as well as the inclusion of explicit treatment guidelines will be necessary to increase the clinical use of measurement-feedback and to possibly reduce disease activity for a larger number of RA patients.

Introduction

Management of patients with Rheumatoid Arthritis (RA) is challenging and poses specific problems. RA has a major impact on function and quality of life. It frequently affects patients in their most productive years, and thus, disability results in a major economic loss. The cornerstone of RA management is the control of disease activity to alleviate pain, maintain function and avoid or slow the rate of joint damage [1]. There is general agreement that rheumatoid disease activity should be controlled as soon as possible, as completely as possible, and that this control should be maintained for as long as possible, consistent with patient safety [2]. Unfortunately, even with the DMARDs and “biologic” medications nowadays available, complete remission or optimal control of disease activity is not achieved in all patients. Further, for individual patients, it cannot be predicted with enough certainty how the course of the disease will develop, if adverse events will occur, and if response or remission will be attained. Thus, we are still challenged to optimise the management of RA patients.

In 1997 the Swiss Clinical Quality Management in RA (SCQM) was introduced [1]. In the Swiss health care system, people have direct access to a rheumatologist, similarly as to

the general practitioner. The SCQM provides a measurement-feedback system with which rheumatologists and their patients can monitor the course of RA disease activity, disability and joint damage [3]. Rheumatologists collect standardised clinical, laboratory and patient data, and send them to a national co-ordination centre, where the data are processed in a computer and a feedback report is returned (Graph 1). With the help of the measurement-feedback system, the individual treatment strategy can be adjusted to "titrate" RA disease activity until remission is reached or disease activity is optimally controlled [1]. Until now, the effectiveness of such a measurement-feedback system in RA has not been the subject of research.

The objective of this trial was to study in RA patients, whether a) the measurement-feedback system is effective in reducing disease activity, and b) the levels of joint damage and disability are consequently maintained or reduced.

Patients and methods

Design

The study was designed as a controlled clinical trial with patients serving as their own controls. A 3-months control period, where disease activity was assessed without feedback, was followed by a 12-months period, where feedback to the rheumatologist was provided. It was hypothesised that the course of disease activity would show a reduction in the feedback period as compared to the control period. As a consequence of reduced disease activity, it was expected that the development of disability and joint damage during the feedback period would stay stable, or even improve.

Recruitment

Rheumatologists from the rheumatology departments of the 5 university hospitals, 6 regional hospitals and from 2 rheumatological practices throughout Switzerland agreed to participate. Patients were recruited during 1997-1998. The rheumatologists asked consecutive RA patients (according to the ACR criteria) to participate in self-assessment of disease activity for a period of 3 months.

Control period

When the patients agreed to participate, they were sent the Rheumatoid Arthritis Disease Activity Index (RADAI) questionnaire on signs and symptoms in RA, once a month, for three months [4]. However, when according to the rheumatologist a change in DMARD

therapy appeared to be necessary, the control period was stopped and the procedure for the feedback period started immediately.

Feedback period

The start of the feedback period was scheduled at the fourth month. The patient was informed by the rheumatologist about participation in the measurement-feedback system. To be included in the study, the patient had to provide written informed consent. In that case, the RADAI from the control period were sent to the co-ordination centre and stored for later analysis. At start, the rheumatologists collected the following data: joint counts, ESR, current medication use, radiographs of the hands and feet (not older than 6 months) and patient assessed disease activity (RADAI), disability (Stanford Health Assessment Questionnaire; HAQ) [5], socio-demographic variables and comorbidities. These data were then sent to the co-ordination centre, where the data were processed and a feedback report was returned within 10 days (Graph 1). For the feedback period, the rheumatologists were advised to monitor disease activity either with every DMARD change or, at the least, every 3 months. For monitoring, disease activity was assessed by the rheumatologist (Disease Activity Score; DAS28) [6], and by the patient (RADAI). An updated feedback report was sent automatically when the rheumatologist sent those data to the co-ordination centre. After 12 months, the last study visit was scheduled, which was identical to the starting visit.

Drop-outs

Patients were excluded from the analysis if: 1) A change in DMARD therapy in the control period was necessary. To avoid influencing the physicians, the patients were excluded after completion of the study. 2) The assessments of the control period were missing.

Measurements

The rheumatologists were provided with standardised information on how to perform the joint counts and how to handle the questionnaires. At the co-ordination centre, the data were checked for completeness and appropriateness before entry. Ambiguities were solved by a telephone call.

The DAS28 was calculated from the results of the 28 swollen joint count, the 28 tender joint count and ESR [6]. The DAS28 ranges virtually from 0 to 10. A DAS28 below 3.2 is regarded as low level disease activity, a DAS28 between 3.2 and 5.1 as moderate, and a

DAS28 larger than 5.1 as high level disease activity [7]. The RADAI is a 5-item patient assessed questionnaire, including arthritis pain, past and current global disease activity, duration of morning stiffness and a tender joint list [4]. The RADAI ranges from 0 to 10, where higher values are indicative for higher levels of RA disease activity. The RADAI has shown to be reliable, valid and responsive for the assessment of disease activity in RA [8-10]. The pain item is an 11-numerical rating scale. The disability index of the HAQ contains 20 questions about difficulties experienced with 8 activities of daily living, and 4 questions about the assistance used to perform these [5]. The HAQ is scored from 0-3 where higher values are indicative for more difficulties when performing activities of daily living. Joint damage was scored from radiographs of the hands and feet by readers blinded for study allocation, using the Ratingen X-ray score [11]. The scoring of the wrist joint is modified by scoring it as a single joint, instead as four joints. The X-ray score ranges from 0 to 160, where higher scores are indicative for more and larger erosions of the joint surface included.

The patient provided socio-demographic information and information about comorbidities on standardised questionnaires [12]. All questionnaires were provided in the language preferred by the patient: German, French, or Italian.

Statistical Analysis

Data from intermediate monitoring visits during the feedback period were not the subject of analysis, as need-driven visits can overestimate the levels of disease activity during this period. Consequently, there are five study time points: 3 in the control period, 2 in the feedback period.

The time course of the RADAI scores in the control period was compared with the course in the feedback period, using a continuous-by-class regression model with random coefficients (intercept and time effect) for patients [13,14]. The procedure thus accounts for repeated measurements on the same subjects. To account for the clustering of patients in rheumatologists, random coefficients (intercept and time effect) for rheumatologists were added.

Changes during the feedback period of disease activity (DAS28), patient perceived pain, disability (HAQ) and joint damage (X-ray score) were analysed using paired t-tests and 95% confidence intervals.

It was hypothesised in advance, that the results could be influenced by the level of disease activity in the control period, and whether measurement-feedback was used during the feedback period. Accordingly, four subgroups were formed: 1) patients with

low disease activity in the control period (RADAI score below the median) and no use of feedback (the rheumatologist had no feedback reports acquired); low disease activity and feedback use (1 or more feedback reports acquired); 3) high disease activity in the control period (RADAI score of median or higher) and no use of feedback; 4) high disease activity and feedback use. The regression analysis and the analysis of before-after differences were repeated as subgroup analysis; differences between subgroups (contrasts) were tested using Scheffe's procedure [15].

To indicate DMARD changes during the feedback period, the medication at start of the feedback period was compared with medication at the end. The information from intermediate visits was not used, to prevent information bias through underreporting in the non-use group.

Data were stored in an Access® 7.0 relational database (Microsoft Corporation, Redmond, USA.) and analysed using SAS® 8.1 (SAS Institute Inc. Cary, NC, USA.). The research protocol was approved by the responsible Swiss medical ethical committee (UREK).

Results

Sample

Forty-eight rheumatologists enrolled 264 patients; 36 patients were drop-outs (Table 1). The RADAI scores and number of dropouts were similar between patients included by private practices, regional hospitals or university hospitals. Reasons for dropping out were: because of DMARD change in the control period (n=33), and because all RADAI in the control period were missing (n=3). No relevant differences on prognostic and outcome variables between dropouts and patients that completed the study were found (Table 1). At the end of the feedback period, 38 patients were lost to follow up. These patients were not regarded as dropouts.

Response and feedback use

The response in the control period varied between 55% and 65% per time point. At the end of the feedback period, 190 (83%) of 228 patients took part in the last study visit. During the feedback period, feedback reports were acquired by the rheumatologists for 142 (62%) patients. Of those, 90 (39%) patients had 1 feedback report, 52 (23%) had 2–5 feedback reports. Of the remaining 86 (38%) patients, the rheumatologists obtained only the report that marked the start of the feedback period.

Was disease activity stable in the control period?

The group mean (sd) RADAI scores in the control period were 3.6 (2.0), 3.7 (2.1) and 3.7 (2.1) respectively. According to the regression model, that is correcting for within-person dependencies, the RADAI scores did not significantly change over time: $\beta_{\text{time}} = -0.008$ (95%CI: -0.12 – 0.11); $p=0.89$ (Table 2).

Was disease activity reduced in the feedback period?

The random time effect for rheumatologists was omitted from the continuous-by-class model, because its associated variance did not significantly differ from 0 (not shown). The time effect in the feedback period was -0.067 per 30 days (Table 2), which corresponds to a mean reduction about 0.8 RADAI points over 12 months. The reduction in RADAI in the feedback period was statistically significant. However, the time effect in the feedback period was not significantly different from the time effect in the control period.

Subgroup analysis

The regression model was subjected to subgroup analysis; the results are shown in the lower part of Table 2. The “level of disease activity” in the control period and “feedback use” in the feedback period were not associated, (Chi-square; $p=0.81$). The difference in time effects was significant in the subgroup with relatively high disease activity in the control period and feedback use in the feedback period ($p=0.02$). The time effect of that subgroup corresponds with a mean reduction in RADAI score over 12 months of more than 3 points. Between both subgroups with high disease activity ($n=111$), the contrast between the difference in time effects of the “use” and “non use” group was at the limit of statistical significance: $p=0.051$.

How did disability and joint damage develop?

During the feedback period, disease activity (DAS28) and pain decreased, disability (HAQ) did not significantly change, but joint damage (X-ray score) increased (Table 3). The subgroup with high disease activity and feedback use, showed significant and favourable changes in DAS28, pain and HAQ (Table 3). Their increase in the X-ray score was comparably small; it was largest in the subgroup with high disease activity and no feedback use. All differences (contrasts) between the subgroups were not statistically significant.

Did changes in medication occur?

The medication changes during the feedback period could only be registered from the patients that were not lost to follow-up and had complete medication information (Table 4). At the start of the feedback period, 32 (14%) of 228 patients had no DMARD therapy, and 20 (9%) patients had a combination therapy of 2 or 3 DMARDs. More than half of the prescriptions (59%) appeared unchanged over the feedback period. Change in DMARD therapy was not significantly associated with the level of disease activity (Chi-square, $p=0.33$), or use of the feedback system (Chi-square, $p=0.50$).

Discussion

According to the results of this study, the use of measurement-feedback was associated with a reduction of RA disease activity in the feedback period as compared to the control period, in patients with high disease activity. The RADAI reduction of that subgroup corresponds to a clinically important difference [10] and was nearly three times larger than the subgroup of patients also with high disease activity, but no feedback use. In concordance, the subgroup with high disease activity and feedback use showed an improvement in the HAQ score and only a small increase in joint damage. In contrast, the subgroup of patients with high disease activity and no feedback use, had no improvement in the HAQ score, and a progression in joint damage that was twice as large. Thus it appears that the measurement-feedback system contributed to a reduction of disease activity in RA patients.

However, it is clear that a measurement-feedback system is not an intervention that causes health effects, but medication may do so. The assumption of a measurement-feedback system for RA is that the system is used to evaluate if disease activity must be better controlled, or to ensure that disease activity is still under control. Then it does not seem to be adequate that changes in DMARD therapy took place in less than half of the patients. Moreover, changes in DMARD therapy were not related to the level of disease activity at baseline or to feedback use. As the study was not designed to include medication strategy in the analysis, and as it is very difficult to judge the appropriateness of specific DMARD management for individual patients, it cannot be concluded that the feedback did not influence decision-making. But the most important conclusion that can be drawn is that not many changes in DMARD therapy occurred, even in patients with high disease activity. Also the use of systematic monitoring with the feedback system was much lower than was originally anticipated on. One of the reasons for the

suboptimal use of the measurement-feedback system may be that available treatment guidelines [16,17] were not explicitly incorporated. It is possible that not all rheumatologists regarded suppression of disease activity as an explicit treatment goal. Also, prescription habits could be influenced by the, meanwhile discarded, “pyramid paradigm” [2,17]. Another reason for the low use of the measurement-feedback in the study may be the local health care system, where rheumatologists merely may have a consulting role for the General Practitioner and see a part of their RA patients probably once yearly. Further, not all rheumatologists are used to outcome measures in clinical practice. Outcome measures are often appraised as “soft” data, unfamiliarity and difficulties with interpretation may lead to uncertainty if, and how, to use all the information [18-21]. But the measurement of RA disease activity and disease consequences has improved substantially, and is within the ability of practicing rheumatologists [22]. The reasons for use and non-use of the measurement-feedback system are currently being studied.

The major advantage of the study design with a control period followed by a measurement-feedback (“intervention”) period is that the patients and rheumatologists are their own controls, and thus are optimally comparable. An important limitation is that it was not possible to blind the participants. However, it was tried to keep the patients naïve towards outcome expectancy, and therefore the primary outcome was patient assessed. A bias from “knowing to be observed” (Hawthorne effect) may have occurred in both the control and feedback periods, and thus may not have introduced differential bias. Due to the fact that the rheumatologists were not blinded, prescription behaviour can theoretically be biased towards socially desired changes in DMARD therapy. However, as the number of patients with changes in DMARD therapy appeared to be low, this may not have played an important role.

It must be noticed that the study of such a complex intervention as measurement-feedback is quite difficult, mainly because: 1) the intervention is mainly addressed to the level of the physician, but important effects are expected on the level of the patient, 2) the intervention is indirect, in the sense that it in itself will not influence disease activity, 3) disease activity is subject to many influencing factors at the same time, e.g. on the levels of patient, treatment, treatment tolerance, social and physical surroundings and prognostic factors of the disease, 4) the effects on outcome are expected to be relatively small, whilst the outcome measures used in RA may not be sensitive enough to detect small but relevant changes.

In primary care, several RCTs were performed on the effectiveness of computerised measurement-feedback systems, with or without guidelines, mainly in patients having hypertension, diabetes, or anticoagulation need [23-29]. From these studies it appears that computerised decision support systems (CDSS) were generally ineffective in changing physician performance or health outcomes, probably because the same kind of difficulties as mentioned above played a role. In arthritis, we could identify two RCTs on measurement-feedback systems, addressing disability [30,31]. Both trials could not find any health gain. The studies used feedback at fixed moments in time, thus the feedback did not systematically coincide with visits or actual patient needs. It is clearly an advantage if relevant information is available at the moment of decision, and decision options are clear. This is the case with measurement-feedback on drug dosing, of which “titration” of disease activity in RA is an example. A systematic review of computerised drug-dosing systems included 5 RCTs regarding out-patient maintenance anticoagulation therapy [32], of which one trial provided evidence that quality of initiation and control of Warfarin treatment was improved by CDSS in comparison to usual care [33]. It follows that it seems justified to conclude that, until now there is no strong evidence of the effectiveness of measurement-feedback systems, computerised drug dosing systems, or computerised guideline implementation systems in several chronic conditions in primary care. However, there is still a strong argument for the adoption of a measurement-feedback system, or some other form of CDSS, in RA. As the primary target of RA treatment is the control of disease activity [17], the treatment has to be individually adjusted depending on the treatment effect and limiting toxicity. For reasons of patient care alone, but especially when using expensive therapies, it is most appropriate to monitor and document medication use, treatment effects and toxicity in the individual [17].

The current study provided limited evidence that a measurement-feedback system is effective in RA, but the system was not intensively used. The measurement-feedback system should be optimized to facilitate its use in clinical practice and its effectiveness should be studied using a carefully designed RCT. For these objectives it should be kept in mind that the primary target of measurement-feedback is the rheumatologist, not the patient. A much more intensive training regarding the use of a measurement-feedback system, the DAS as target measure, and the inclusion of explicit treatment guidelines will be necessary to increase the clinical use of measurement-feedback and to possibly reduce disease activity for a larger number of patients.

Acknowledgements

We are indebted to the SCQM Members as well as to all other rheumatologists that took part in the study, especially Drs: W Schmid (Aarau), P Hasler (Basel), T Lehmann and J Lory (Berne), G-A Davoine and F Perret (Geneva), J Dudler (Lausanne), T Stoll (Schinznach), U Schlumpf (Lucerne), R Rüdts and B Christen (Winterthur).

The members of the Swiss Clinical Quality Management in Rheumatoid Arthritis (SCQM) are: I Büchler, B Christen, A-M Chamot, G Dalvit, A Forster, P-A Guerne, P Hasler, T Langenegger, BA Michel, L Schmid, H Schwarz, M Seitz, A So, R Theiler, D Van Linthoudt, P Villiger.

The Swiss Clinical Quality Management programme and the current trial were supported by grants from the Swiss Academy for Medical Sciences (SAMW), the Swiss Health Authorities (BAG) and the Böni Foundation.

The authors wish to thank Leanne Pobjoy for her help in preparing the manuscript.

References

1. Stucki G, Sangha O. Clinical quality management: putting the pieces together. *Arthritis Care Res.* 1996;9(5):405-12.
2. Moreland LW, Russell AS, Paulus HE. Management of rheumatoid arthritis: the historical context. *J Rheumatol.* 2001;28:1431-52.
3. Uitz E, Fransen J, Langenegger T, Stucki G. Clinical quality management in rheumatoid arthritis: putting theory into practice. *Swiss Clinical Quality Management in Rheumatoid Arthritis. Rheumatology (Oxford).* 2000;39(5):542-9.
4. Stucki G, Liang MH, Stucki S, Bruhlmann P, Michel BA. A self-administered rheumatoid arthritis disease activity index (RADAI) for epidemiologic research. Psychometric properties and correlation with parameters of disease activity. *Arthritis Rheum* 1995;38:795-8.
5. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
6. Prevoo MLL, Van 't Hof MA, Kuper HH, Van Leeuwen MA, Van de Putte LBA, Van Riel PLCM. Modified disease activity scores that include twenty-eight joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum.* 1995;38:44-8.
7. Van Gestel AM, Prevoo MLL, Van 't Hof MA, Van Rijswijk MH, Van de Putte LBA, Van Riel PLCM. Development and validation of the european league against rheumatism response criteria for rheumatoid arthritis. *Arthritis Rheum* 1998;41:1845-50.
8. Feasibility and validity of the RADAI, a self-administered Rheumatoid Arthritis Disease Activity Index. Fransen J, Langenegger T, Michel BA, Stucki G. *Rheumatology Oxford.* 2000;39:321-7.
9. Responsiveness of the self-assessed rheumatoid arthritis disease activity index to a flare of disease activity. Fransen J, Häuselmann H, Michel B, Caravatti M, Stucki G. *Arthritis Rheum.* 2001;44(1):53-60.
10. Reliability and responsiveness of the RADAI, a self-assessed rheumatoid arthritis disease activity index. Fransen J, Forster A, Uebelhart D, Michel BA. *Ann Rheum Dis.* 2001;60(Suppl 1):345.
11. Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. *J Rheumatol* 1998;25(11):2094-107.
12. Sangha O, Stucki G, Fossel AH, Katz JN. A simplified method to assess comorbidity in clinical and health service research of rheumatic diseases. *Arthritis Rheum* 1995;38(S):177, abstract 148.
13. SAS/STAT® User's Guide, Version 6 Fourth Edition, Volume 2. 1990 SAS Institute, Cary NC, USA.
14. Verbeke G, Molenberghs G. Linear mixed models in practice. A SAS-oriented approach. Springer Verlag, New York 1997.
15. Kleinbaum DG, Kupper LL, Muller KE. Applied regression analysis and other multivariable methods. Boston: PWS-KENT, 1988.
16. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. Guidelines for the management of rheumatoid arthritis. *Arthritis Rheum.* 1996;39(5):713-22.

17. Wolfe F, Cush JJ, O'Dell JR, Kavanaugh A, Kremer JM, Lane NE et al. Consensus recommendations for the assessment and treatment of rheumatoid arthritis. *J Rheumatol.* 2001;28:1413-30.
18. Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care* 1989;27(suppl):S254-68.
19. Deyo RA, Carter WB. Strategies for improving and expanding the application of health status measures in clinical settings: a researcher-developer viewpoint. *Med Care* 1992;30(suppl):S176-186.
20. Golden WE. Health status measurement. Implementation strategies. *Med Care* 1992;30(suppl):S187-195.
21. Patrick DL: Strategies for improving and expanding the application of health status measures in clinical settings. Discussion. *Med Care* 1992;30(suppl):S198.
22. Pincus T. Documenting quality management in rheumatic disease: are patient questionnaires the best (and only) method? *Arthritis Care Res.* 1996;9(5):339-48.
23. Goldberg HI, Wagner EH, Fihn SD, Martin DP, Horowitz CR, Christensen DB et al. A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? *J Qual Improvement.* 1998 ;24(3) :130-42.
24. Montgomery AA, Fahey T. A systematic review of the use of computers in the management of hypertension. *J Epidemiol Community Health.* 1998;52:520-5.
25. Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. *BMJ.* 2000;320:686-90.
26. Hetlevik I, Holmen J, Kruger O. Implementing clinical guidelines in the treatment of hypertension in general practice. *Scand J Prim Health Care.* 1999;17(1):35-40.
27. Hetlevik I, Holmen J, Kruger O, Kristensen P, Iversen H, Furuseth K. Implementing clinical guidelines in the treatment of diabetes mellitus in general practice. *Int J Technol Assess Health Care.* 2000;16(1):210-27.
28. Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *JAMIA* 1999;6:104-114.
29. Johnston ME, Langton KB, Haynes B, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med.* 1994;120:135-42.
30. Rubenstein LV, Calkins DR, Young RT, Cleary PD, Fink A, Kosecoff J et al. Improving patient function: a randomized trial of functional disability screening. *Ann Internal Medicine.* 1989;111(10):836-42.
31. Kazis LE, Meenan RF, Anderson JJ, Swift MA. The clinical utility of health status information in rheumatoid arthritis (RA) patients: a controlled trial. *Arthritis Rheum.* 1987;30(4 Suppl):S12.
32. Hunt DL, Haynes B, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *JAMA* 1998;280(15):1339-46.
33. Vadher B, Patterson DL, Leaning M. Evaluation of a decision support system for initiation and control of oral anticoagulation in a randomised trial. *BMJ.* 1997;314(7089):1252-6.

Table 1 Population characteristics at baseline

	Included		Dropouts		p
Patients, n=(%)	228		36		
Female, n= (%)	156	(68%)	32	(89%)	0.01
Age years, mean (sd)	59	(13)	60	(15)	0.81
TS years, median (IQR)	12	(6-18)	9	(4-15)	0.15
TD years, median (IQR)	11	(4-17)	8	(4-14)	0.30
RF +, n= (%)	161	(80%)	27	(84%)	0.57
ANA +, n= (%)	73	(38%)	7	(23%)	0.12
Pain, median (IQR)	3	(1-5)	4	(2-6)	0.32
RADAI, mean (sd)	3.5	(2.0)	3.8	(1.8)	0.47
DAS28, mean (sd)	4.1	(1.5)	4.4	(1.2)	0.19
HAQ, median (IQR)	1.0	(0.4-1.8)	1.1	(0.5-1.5)	0.49
X-ray score, median (IQR)	3	(0-12)	1	(0-21)	0.71

Values are mean (sd); median (interquartile-range); number (column percentage).
 TS: time since symptom onset; TD: time since diagnosis; RF+: rheumatic Factor positivity; ANA+: Anti Nuclear Antibody positivity. RADAI: Rheumatoid Arthritis Disease Activity Index; DAS28: Disease Activity Index; HAQ: Health Assessment Questionnaire; X-ray: Ratingen X-ray score.

Graph 1 An example of the feedback to the rheumatologists.

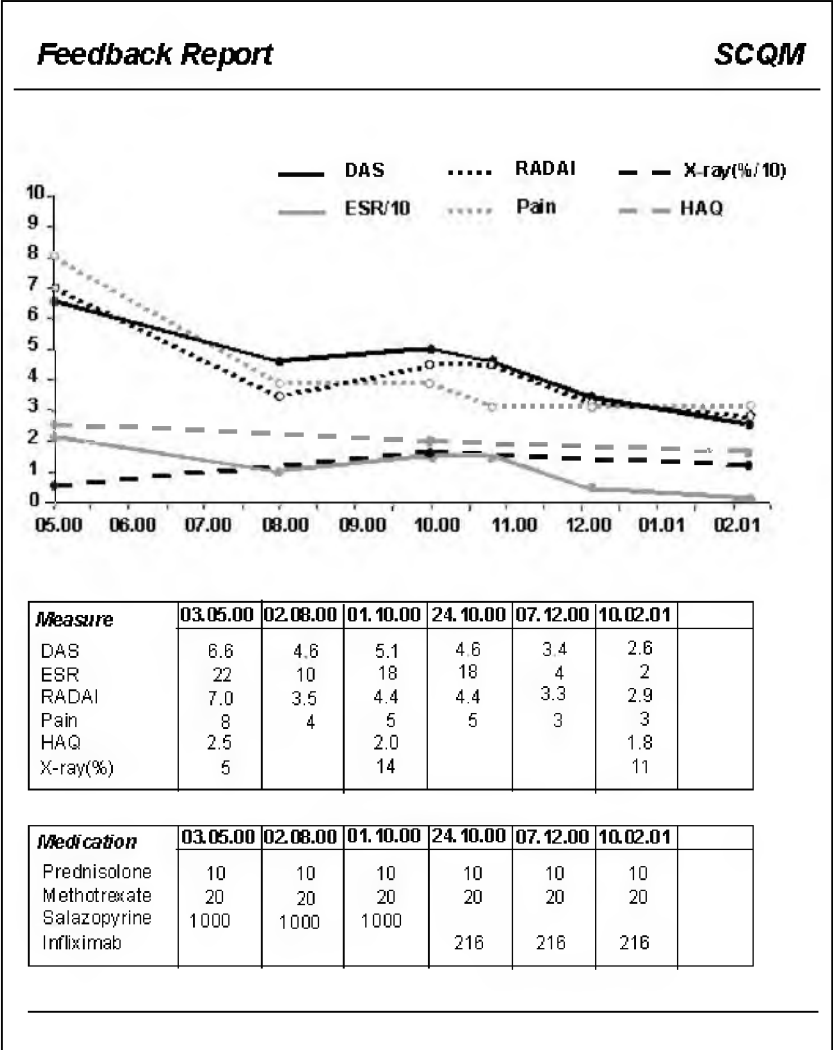


Table 2 Difference between control- and feedback period of the course of the RADAI over time.

	n=	Intercept β_0	95%CI	Control Period Time Effect $\beta(C)$	95%CI	Feedback Period Time Effect $\beta(F)$	95%CI	Difference $\beta(C) - \beta(F)$	95%CI
Total	228	3.27 ***	(2.91 – 3.63)	-0.008 ^{ns}	(-0.12 – 0.11)	-0.067 ***	(-0.095 – -0.039)	-0.059 ^{ns}	(-0.177 – 0.058)
Subgroups									
1) Low DA, no use	45	2.07 ***	(1.67 – 2.47)	-0.033 ^{ns}	(-0.24 – 0.17)	-0.015 ^{ns}	(-0.057 – 0.026)	0.017 ^{ns}	(-0.18 – 0.22)
2) Low DA, use	72	2.08 ***	(1.75 – 2.40)	-0.016 ^{ns}	(-0.25 – 0.22)	-0.011 ^{ns}	(-0.071 – 0.049)	0.0054 ^{ns}	(-0.23 – 0.24)
3) High DA, no use	41	4.76 ***	(4.33 – 5.19)	-0.003 ^{ns}	(-0.21 – 0.20)	-0.10 ***	(-0.15 – -0.053)	-0.097 ^{ns}	(-0.29 – 0.10)
4) High DA, use	70	4.94 ***	(4.59 – 5.30)	0.076 ^{ns}	(-0.15 – 0.30)	-0.19 ***	(-0.26 – -0.12)	-0.27 *	(-0.49 – -0.041)

Ns $p > 0.05$; * $p \leq 0.05$; ** $p < 0.001$; *** $p < 0.0001$. RADAI: Rheumatoid Arthritis Disease Activity Index. DA: disease activity. See methods for subgroup definitions.

Table 3 Subgroup analysis of changes in estimators of disease activity, disability and joint damage in the feedback period.

	n=	Δ DAS28		Δ Pain		Δ HAQ		Δ X-Ray Score	
		Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Total	190	-0.3 ***	(-0.5 – -0.2)	-0.6 **	(-1.0 – -0.3)	-0.05	(-0.11 – 0.01)	6.7 ***	(4.6 – 8.8)
Subgroups									
1) Low DA, no use	38	-0.2	(-0.6 – 0.2)	-0.2	(-0.9 – 0.6)	0.05 **	(-0.09 – 0.20)	6.2 **	(2.8 – 9.7)
2) Low DA, use	64	-0.3 *	(-0.6 – 0.005)	-0.3	(-0.8 – 0.3)	-0.01	(-0.10 – 0.08)	5.5 **	(2.4 – 8.6)
3) High DA, no use	30	-0.4	(-1.0 – 0.2)	-1.4 *	(-2.5 – -0.3)	-0.06	(-0.21 – 0.07)	13.1 *	(4.7 – 21.4)
4) High DA, use	58	-0.4 **	(-0.7 – -0.1)	-0.9 *	(-1.7 – -0.2)	-0.14 *	(-0.3 – -0.04)	5.4 *	(1.0 – 9.8)

* $p \leq 0.05$; ** $p < 0.001$; *** $p < 0.0001$. DAS28=Disease Activity Score, HAQ=Disability Index of the Health Assessment Questionnaire, X-ray: Ratingen X-ray score. DA: disease activity. See methods for subgroup definitions.

Table 4 Changes in DMARD medication at one-year follow-up.

Subgroup	Follow-up n=	Non-missing n=	DMARD changes							
			No change		Dose reduced		Dose increased		Changed	
1) Low DA, no use	38	33	22	(67%)	2	(6%)	5	(15%)	4	(12%)
2) Low DA, use	64	60	37	(62%)	8	(13%)	10	(17%)	5	(8%)
3) High DA, no use	30	23	9	(39%)	5	(22%)	5	(22%)	4	(17%)
4) High DA, use	58	53	32	(60%)	5	(10%)	7	(13%)	9	(17%)
Total	190	169	100	(59%)	20	(12%)	27	(16%)	22	(13%)

The percentages were calculated over the non-missing numbers per subgroup; in 89% of the 190 patients at follow-up, medication information was complete (non-missing).

Chapter 7

Influence of guideline adherence on outcome in a randomized controlled trial on the efficacy of methotrexate with folate supplementation in rheumatoid arthritis.

Jaap Fransen¹, Roland Laan¹, Mart van der Laar², Tom Huizinga³ and Piet van Riel¹

¹ Department of Rheumatology, University Medical Centre Nijmegen, The Netherlands.

² Department of Rheumatology, Medisch Spectrum Twente, Enschede, The Netherlands.

³ Department of Rheumatology, Leiden University Medical Centre, The Netherlands.

Abstract

Objective To study the influence of rheumatologists' guideline adherence on efficacy and toxicity of treatment with methotrexate (MTX) in a 48 week, randomized controlled trial of MTX with supplementation of folates or placebo in rheumatoid arthritis.

Methods To reach an optimal dose of MTX, guidelines based on the disease activity score (DAS) and the occurrence of adverse events were applied. MTX was started with 7.5 mg/week and raised every 6 weeks by 2.5 mg/week, until a good response was reached or adverse events occurred. The course of the DAS was analyzed using generalized estimating equations; the occurrence of adverse events was analyzed with survival analysis corrected for time dependency.

Results In 51% of the 411 study patients the guidelines were always followed. In 25% of patients, non-adherence lead to lower doses and in 24% to higher doses of MTX than the guidelines had proposed. In the adherence group, the reduction of the DAS was significantly larger (mean 0.4; $p=0.0085$) compared with the 'low dose' group; the 'high dose' and adherence groups did not differ (mean -0.07; $p=0.64$). The dropout due to severe adverse events appeared not to be different between the three groups.

Conclusion The effect of prescribing MTX doses higher than the guidelines proposed seemed not to be more beneficial than guideline adherence, whereas the prescription of MTX doses lower than the guidelines proposed reduced efficacy without clear beneficial effect on toxicity. Guidelines assistive in determining the optimal efficacious MTX or DMARD dose may be useful in daily clinical practice.

Introduction

The most important aim in the treatment of rheumatoid arthritis (RA) is the reduction and control of rheumatoid inflammation and the prevention of irreversible joint damage [1]. As a treatment principle, anti-rheumatic medication should be used in doses that are sufficiently high to reduce inflammation, unless limiting toxicity is reached [1]. The anti-rheumatic drugs nowadays available have the potential to act relatively fast. This gives the rheumatologist better opportunities to "titrate" DMARD therapy to get RA disease activity under control [2]. However, in practice it may be difficult to find the optimal effective anti-rheumatic drug dose for an individual RA patient. An important reason is the difficulty in the assessment and the judgement of rheumatoid inflammation. Therefore, a combination of standardized systematic evaluation of rheumatoid inflammation and guidelines to assist in its judgement may be helpful in the treatment of

RA patients [3]. Generally, clinical guidelines aim at the improvement of health care. However, the question is justified if guidelines can be effective in changing physician performance and can eventually lead to improved health outcome [4]. For the case of RA, we had the opportunity to study the influence of guideline adherence on outcome in a clinical trial [5]. The objective was to study whether rheumatologists' guideline adherence had an influence on efficacy and toxicity of treatment with methotrexate (MTX), in a randomized controlled trial of MTX with supplementation of folates or placebo in RA.

Patients and methods

Design

The study was performed by post-hoc analysis of the data from a 48-week, multicenter, randomized, double blind and placebo controlled trial (RCT), on the effect of supplementation of folic or folinic acid on toxicity and efficacy of MTX treatment in RA [5]. In the RCT, eligible patients with RA according to the ACR criteria were randomly allocated to receive MTX with addition of placebo or folic- or folinic acid. Rheumatology units from university and regional hospitals were participating. The same study nurse monitored the patients for adverse events every 3 weeks. Every 6 weeks, the patients were additionally monitored for effectiveness and visited the treating rheumatologist. In case of adverse events, visits could be more frequent. Complete information on patients and methods is given in the original publication [5]. The study protocol was approved by the ethical committees of all participating hospitals.

Guidelines

MTX was taken once weekly, with a starting dose of 7.5 mg/week and a maximal dose of 25 mg/week. To realize an optimal dose of MTX in the trial, guidelines for increasing and reducing the dose were established. Every 6 weeks, the MTX dose could be increased with 2.5 mg/week, until a good response was reached. The standardized response criteria were a preliminary version of the EULAR criteria, and compared the actual level of the Disease Activity Score (DAS) with the level at start of treatment [6,7]. In the presence of adverse events, the MTX dose was decreased (mild adverse events), or the use of MTX was temporarily (moderate adverse events) or finally stopped (severe adverse events) [5]. The guidelines were applied by a medical study coordinator who calculated the DAS and judged efficacy and toxicity based on the results of the monitoring. The treating rheumatologist was advised regarding the MTX dose, by phone.

The treating rheumatologists were allowed to deviate from the dose proposed by the guidelines. The MTX dose that was finally prescribed was registered.

Measures

The DAS was calculated using the Ritchie articular index (RAI), a swollen joint count, erythrocyte sedimentation rate (ESR) and general health [6]. The RAI was calculated according to the grading and accumulation described by Ritchie et al., and ranged from 0 - 78 [8]. The swollen joint count ranged from 0 - 44. General health (GH) and pain were rated on 100 mm. visual analogue scales (VAS). Patient- and physician global disease activity were rated on scales ranging from 0 (no activity) to 5 (very severe activity). The importance of possible adverse events was rated by the patients, and used in the calculation of toxicity indices according to Fries and the modification by Felson [9,10].

Outcomes

For this study, the course of the DAS over time was regarded as the primary outcome for efficacy. Primary outcome for toxicity was the occurrence of severe adverse events, consequently leading to the definitive stop of MTX. Secondary outcomes were the reduction in the DAS at 48 weeks, the response at 48 weeks, the time needed to reach a good response, the number of observations in good response, trial dropout for any reason, and the toxicity indices at 48 weeks.

Guideline adherence

The adherence of the rheumatologists to the guidelines as applied by the study coordinator was judged after the study had been completed. Adherence was determined from the database, by comparing the prescribed MTX dose with the dose proposed by the guidelines. If all MTX prescriptions for an individual patient were in congruence with the guidelines, this was determined a case of full adherence (FA). A case of non-adherence (NA) was determined if one or more decisions were not in agreement with the guidelines. In case of non-adherence, the dose deviation from the guidelines was calculated. If the dose deviation was positive, the patient had received more MTX than the guidelines proposed, then the patient was classified as NA+; if the dose deviation was negative the patient was classified as NA-.

Statistical analysis

Baseline differences between the three adherence groups (FA, NA-, NA+) were analyzed with the one-way ANOVA, the Kruskal-Wallis test or the Chi-square test, where appropriate. The result of non-adherence on MTX dose was analyzed by testing the between-group differences of the cumulative dose of MTX at 48 weeks, using one-way ANOVA and contrasts.

The course of the DAS over time was analyzed using a non-linear regression model with random coefficients for patient, suited for repeated measurements (generalized estimating equations) [11]. Differences between the adherence groups were tested using contrasts. The dropout due to severe adverse events was analyzed using survival analysis, accounting for the fact that the final adherence status was not known at start of the trial [12]. If that would be neglected, it would seem that guideline adherence causes early drop-out, while in fact early drop-out may “cause” adherence by preventing from committing non-adherence. Differences in drug survival between the adherence groups were analyzed using Chi-square tests [12].

The survival analysis was repeated with dropout for any reason (patient wish, protocol violation, inefficacy, and severe adverse events). Differences between the three adherence groups in DAS reduction and response at 48 weeks, time-to-good-response, number of observations in good response, and in the toxicity indices, were analyzed using a Chi-square test, one-way ANOVA, or a Kruskal-Wallis test, where appropriate. A Bonferroni procedure was applied when comparing the non-adherence groups with the full adherence group, by dividing $\alpha=0.05$ by 2, giving a border of statistical significance of 0.025. The data were analyzed using SAS 8.1 (SAS Institute Inc. Cary, USA).

Results

Guideline Adherence

In 208 (51%) of the 411 patients included in the RCT, all decisions on MTX dosage taken by the treating rheumatologist were in agreement with the guidelines as applied by the study coordinator. In contrast, in 203 (49%) patients a total of 390 decisions were not in agreement with the guidelines. The decisions involved in non-adherence are listed in Table 1. The deviations from the guidelines were generally not larger than one dose step in the protocol: 2.5 mg/week. As a result, individual patients could have received less MTX (NA-, n=102) or more MTX (NA+, n=101) than the guidelines had proposed for them.

Baseline

In Table 2, it is shown that both forms of non-adherence (NA- and NA+) were equally divided among the three original treatment arms of the trial. At baseline, differences existed in gender (more female in NA+) and disease duration (NA+ shorter). A small and non-significant difference appeared in baseline values of the DAS.

MTX Dose

In Graph 1, the deviations of the cumulative MTX dose of both non-adherence groups as compared with the adherence group are shown. After week 24, the median cumulative dose of the NA- group did not increase as much as in the NA+ and FA groups. At the end of the trial, the resulting cumulative MTX dose of the NA- group (median 615 mg) was significantly lower ($p=0.007$) than the cumulative MTX dose of the FA group (median 650 mg). It can also be seen in Graph 1 that after week 18 the median cumulative MTX dose increased in the NA+ group as compared with the FA group. After week 36 this difference nearly diminished, by dose increases in the FA group. At week 48, the cumulative dose of the NA+ group (median 652.5 mg) was comparable ($p=0.93$) to the FA group (median 650 mg).

Efficacy

In Graph 2, the course of the DAS over time of the three adherence groups is shown. There was a small but significant overall difference in the course of the DAS between the three groups ($p=0.028$). When testing contrasts, the reduction of the DAS over time in the NA- group was significantly smaller than in the FA group ($p=0.0085$), whereas the course of the DAS of the NA+ group did not differ from the FA group ($p=0.64$). Correction for the level of DAS at baseline did not change the results (not shown). At week 48, the decrease in the DAS was larger in the NA+ and FA groups than in the NA- group (Table 3). Furthermore, the NA+ and FA groups had a larger fraction of good responders, a shorter time to good response and a longer time in response, than the NA- group (Table 3).

Toxicity

The dropout due to severe adverse events was not different between NA- group and the FA group (Chi-square value=0.19 (1 DF); $p>0.60$) and between the NA+ and FA groups (Chi-square value=0.14 (1 DF); $p>0.70$). The survival curves are shown in Graph 3. When analyzing dropout for any reason, there were again no significant differences

between the adherence groups (not shown). At 48 weeks, no significant differences in the Toxicity Indices appeared between the three adherence groups (Table 3).

Discussion

To our knowledge, this is the first study to demonstrate that adherence to guidelines for drug dosing can make a difference in efficacy. The prescription of MTX doses lower than the guidelines proposed reduced efficacy and had no clear beneficial effect on toxicity. The effect of prescribing higher doses than the guidelines proposed did not seem to be more beneficial than full adherence to the guidelines, presumably because the resulting cumulative MTX doses were comparable. Thus, in case of guidelines using the DAS for MTX dosing in RA, strict adherence to the guidelines appears to be unproblematic and moreover, more beneficial, than the use of a more loosely “individualized” guideline interpretation.

As it was strived to reach optimal guideline adherence in the trial, the differences in MTX dose between the three adherence groups were not very large, thus large differences in outcome cannot be expected. In the original trial, no significant differences in the course of the DAS between the three treatment arms were found [5]. It is important to notice that (non-) adherence to the guidelines was equally divided among the three original trial arms. Also, the rheumatologists were allowed to deviate from the trial guidelines for MTX dosing. Thus, we feel there is no reason to believe that the results are biased due to systematic differences in study medication or socially desirable dose registration. The most frequent reasons to deviate from the guidelines were on patient wish, to prevent possible adverse reactions, or to induce a response sooner. The main limitation of this study is, that patients were not randomized to the use of guidelines or not to use guidelines. Thus, to be able to answer the question whether clinical guidelines on drug dosing indeed can be used to influence physician performance and health outcomes in RA patients, it is necessary to use an RCT design, comparing the use of clinical guidelines versus “usual care”.

For the use of guidelines, it should be noted that clinical guidelines are not intended to replace clinical judgement, and also are not necessarily the only way in which a patient can be treated [3]. But the contribution of guidelines to clinical practice is that they provide an analytical framework for evaluation and treatment [3]. While it is not difficult to subscribe these advantages, multiple barriers can limit physician guideline adherence and translation into improved patient outcomes. These barriers can be associated with the patient, the guideline, and the practice environment, but can also be physician-bound

factors as: lack of awareness, lack of familiarity, lack of agreement, lack of self-efficacy, lack of outcome expectancy, and inertia of previous practice [13]. These barriers may in part explain why to date there is not much evidence to show that guidelines on medication prescription can change physician behavior and consequently influence health outcomes. Several RCTs have been performed for clinical issues similar to the titration of DMARDs for controlling disease activity in RA, especially anti-coagulation therapy [14-18], hypertension [19-21] and diabetes [22]. In RA, 2 studies incorporated guidelines and monitoring for DMARD dose titration, pointing to the feasibility of a combination of monitoring and guidelines on drug dosing [5,23]. In the RCT on the effects of the addition of folates or placebo to MTX that was used for this study, guidelines using the DAS and explicit response criteria were established to individually increase the dose of MTX using a step-up protocol [5]. In a follow-up study of a cohort of RA patients on anti-TNF α , the advantages of tailoring anti-TNF α treatment compared to the “one size fits all” dosing scheme were demonstrated using a step-down protocol [23]. It was shown that the total amount of anti-TNF α given could be reduced by 67%, while the level of disease activity was maintained and no patients dropped out due to persistent worsening of disease activity.

For use in daily clinical practice, a monitoring system has been developed for RA [24]. It allows the systematic registration and (graphical) representation of several parameters of disease activity, disability and joint damage, together with medication data. The course of disease activity over time can be followed and compared with guidelines or another external reference. How the DAS can be used for the titration of DMARD dose in the suppression of disease activity in individual patients in daily clinical practice is explained by Van Riel [25]. Because the current study was not comparing the use of guidelines versus no guidelines, it cannot be said how effective those guidelines would be in clinical practice, which should be studied using a RCT design. However, already within the framework of the RCT on folate supplementation to MTX [5], a beneficial effect of guideline adherence appeared. As it can be expected that practice variation in prescribing MTX dose in daily clinical practice is much larger than in this study, clinical guidelines that are sufficiently adhered to may have a larger influence on efficacy, and perhaps on toxicity, in daily clinical practice.

Acknowledgements

The RCT on the effect of folic or folinic supplementation on the toxicity and efficacy of methotrexate in rheumatoid arthritis was funded by Ontwikkelingsgeneeskunde, Ziekenfondsraad, The Netherlands (grant number 95-016).

We wish to thank Erik Brummelkamp for his assistance in handling the database.

References

1. Wolfe F, Cush JJ, O'Dell JR, Kavanaugh A, Kremer JM, Lane NE, et al. Consensus recommendations for the assessment and treatment of rheumatoid arthritis. *J Rheumatol* 2001;28(6):1423-30.
2. Van Gestel AM, Van Riel PLCM. Evaluation of early rheumatoid arthritis disease activity and outcome. *Baillieres Clin Rheumatol* 1997;11(1):49-63.
3. Scott DL. Clinical guidelines for management. *Baillieres Clin Rheumatol* 1997;11(1):157-179.
4. Cibere J, Sibley JT, Haga M. Rheumatologists' adherence to guidelines for misoprostol use in patients at high risk for nonsteroidal antiinflammatory drug gastropathy. *J Rheumatol* 2002;29(2):339-46.
5. Van Ede EA, Laan RFJM, Rood MJ, Huizinga TWJ, Van de Laar MAFJ, Van Denderen CJ, et al. Effect of folic or folinic supplementation on the toxicity and efficacy of methotrexate in rheumatoid arthritis. A forty-eight-week, multicenter, randomized, double-blind, placebo-controlled study. *Arthritis Rheum* 2001;44(7):1515-24.
6. Van der Heijde DMFM, Van 't Hof MA, Van Riel PLCM, Van de Putte LBA. Development of a disease activity score based on judgement in clinical practice by rheumatologists. *J Rheumatol* 1993;20(3):579-81.
7. Van Gestel AM, Prevoo MLL, Van 't Hof MA, Van Rijswijk MH, Van de Putte LBA, Van Riel PLCM. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism criteria. *Arthritis Rheum* 1996;39(1):34-40.
8. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakos TG, Grieveson P, et al. *Q J Med* 1968;37(147):393-406.
9. Fries JF, Spitz PW, Williams CA, Bloch DA, Sing G, Hubert HB. A toxicity index for comparison of side effects among different drugs. *Arthritis Rheum* 1990;33(1):121-30.
10. Felson DT, Anderson JJ, Meenan RF. Use of short-term efficacy/toxicity tradeoffs to select second-line drugs in rheumatoid arthritis. A metaanalysis of published clinical trials. *Arthritis Rheum* 1992;35(10):1117-25.
11. Verbeke G, Molenberghs G. Linear mixed models in practice. A SAS-oriented approach. New York: Springer Verlag; 1997.
12. Simon R, Makuch RW. A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias. *Stat Med* 1984;3(1):35-44.
13. Cabana MD, Rushton JL, Rush AJ. Implementing practice guidelines for depression: Applying a new framework to an old problem. *Gen Hosp Psychiatry* 2002;24(1):35-42.
14. Fitzmaurice DA, Hobbs FD, Delaney BC, Wilson S, McManus R. Review of computerized decision support systems for oral anticoagulation management. *Br J Haematol* 1998;102(4):907-9.

15. Fitzmaurice DA, Hobbs FD, Murray ET, Bradley CP, Holder R. Evaluation of computerized decision support for oral anticoagulation management based in primary care. *Br J Gen Pract* 1996;46(410):533-5.
16. Johnston ME, Langton KB, Haynes B, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med* 1994;120:135-42.
17. Hunt DL, Haynes B, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *JAMA* 1998;280(15):1339-46.
18. Vadher B, Patterson DL, Leaning M. Evaluation of a decision support system for initiation and control of oral anticoagulation in a randomised trial. *BMJ* 1997;314(7089):1252-6.
19. Hetlevik I, Holmen J, Kruger O. Implementing clinical guidelines in the treatment of hypertension in general practice. *Scand J Prim Health Care* 1999;17(1):35-40.
20. Montgomery AA, Fahey T. A systematic review of the use of computers in the management of hypertension. *J Epidemiol Community Health* 1998;52:520-5.
21. Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomised controlled trial. *BMJ* 2000;320:686-90.
22. Hetlevik I, Holmen J, Kruger O, Kristensen P, Iversen H, Furuseth K. Implementing clinical guidelines in the treatment of diabetes mellitus in general practice. *Int J Technol Assess Health Care*. 2000;16(1):210-27
23. Den Broeder AA, Creemers MCW, Van Gestel AM, Van Riel PLCM. Dose titration using the disease activity score (DAS28) in rheumatoid arthritis patients treated with anti-TNF α . *Rheumatology (Oxford)* 2002;41(6):638-42.
24. Uitz E, Fransen J, Langenegger T, Stucki G. Clinical quality management in rheumatoid arthritis: putting theory into practice. *Swiss Clinical Quality Management in Rheumatoid Arthritis. Rheumatology (Oxford)* 2000;39(5):542-9.
25. Van Riel PLCM, Schumacher HR. How does one assess early rheumatoid arthritis in daily clinical practice? *Best Pract Res clin Rheumatol* 2001;15(1):67-76.

Table 1 Decisions involved in non adherence to the guidelines.

	Non Adherence		Total
	NA -	NA +	
Patients (n)	102	101	203
Decisions (No.)	201	189	390
Specification of decisions			
Restart with less than last tolerated dose	6		6
Restart with more than last tolerated dose	3	22	25
No increase, though no good response	132	26	158
No increase, for other reasons	15	17	32
Increased, though good response	1	43	44
Increased with 5.0 mg/week		15	15
Increased with 7.5 mg/week		3	3
Preterm increase	30	45	75
Dose above 25 mg/week		6	6
Decrease below last tolerated dose	8	1	9
Decrease too small		1	1
Decrease by patient initiative	1		1
No temporarily stop, though moderate adverse event		6	6
No decrease, though mild adverse event	3	3	6
Unjustified temporarily stop	2		2

NA - : Patients with a cumulative MTX dose at the end of the trial period that was lower than the guidelines proposed; NA + : Patients with a cumulative MTX dose at the end of the trial period that was higher than the guidelines proposed. No. : Number.

Table 2 Population characteristics at baseline.

	Non Adherence		Adherence		p	Total	
	NA -	NA +	FA				
N	102	101	208			411	
MTX + folinic acid	41 (40%)	35 (35%)	65 (31%)		0.53	141 (34%)	
MTX + folic acid	28 (27%)	31 (31%)	74 (36%)			133 (32%)	
MTX + placebo	33 (32%)	35 (35%)	69 (33%)			137 (33%)	
Female (n)	62 (61%)	75 (74%)	153 (74%)		0.04	290 (71%)	
RF+ (n)	85 (88%)	78 (80%)	155 (78%)		0.17	318 (77%)	
Age (Years)	57 (14)	53 (14)	58 (11)		0.07	56 (13)	
Disease duration (Mt.)	47 (12-132)	27 (12-72)	51 (12-130)		0.04	45 (12-117)	
Disease Activity Score	4.7 (1.3)	4.9 (0.9)	5.0 (1.1)		0.07	4.9 (1.1)	
Ritchie score	17 (10-26)	18 (12-22)	19 (12-28)		0.18	18 (12-26)	
Swollen joint count	17 (12-22)	16 (11-21)	17 (12-23)		0.38	17 (12-22)	
General health	47 (21)	52 (20)	48 (19)		0.13	49 (20)	
Pain score	48 (20)	54 (19)	49 (21)		0.14	50 (20)	
Patient global	3 (3-4)	3 (3-3)	3 (3-4)		0.42	3 (3-4)	
Physician global	3 (3-4)	3 (3-4)	3 (3-4)		0.79	3 (3-4)	

NA - : Patients with a net lower MTX dose than the guidelines proposed; NA + : Patients with a net higher MTX dose than the guidelines proposed; FA : Patients with full guideline adherence.

Mt. : months; Patient global: Patient's global assessment of disease activity; Physician global: Physician's global assessment of disease activity.

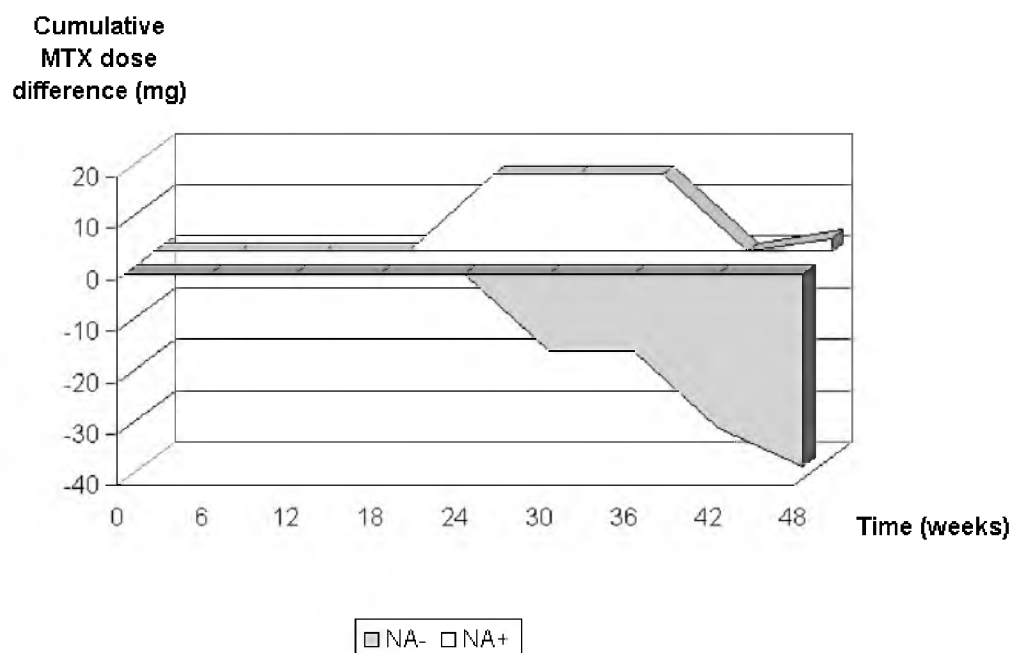
Values are counts (column percentage); mean (standard deviation); median (interquartile-range).

Table 3 Disease activity and toxicity at week 48.

	NA -	FA	NA +	p	Contrasts	
					NA- vs. FA	NA+ vs. FA
N	97	180	98			
DAS	3.2 (1.2)	3.2 (1.4)	3.1 (1.1)	0.61	--	--
Decrease in DAS	-1.4 (1.2)	-1.8 (1.3)	-1.8 (1.2)	0.05	0.020	0.80
DAS response none	5 (5%)	16 (8%)	4 (4%)			
moderate	57 (56%)	70 (33%)	43 (42%)	0.02	0.011	0.30
good	35 (34%)	92 (44%)	50 (49%)			
Time-to-response* (Weeks)	27 (12)	24 (11)	21 (12)	0.0062	0.14	0.031
Observations in response* (No.)	2 (1-4)	4 (2-5)	4 (2-6)	0.0003	0.0003	0.57
Toxicity index Fries	10.9 (4.3-20.4)	8.8 (2.4-19.2)	11.4 (5.3-19.6)	0.30	--	--
Toxicity index Felson	15.4 (13.4-20.7)	10.9 (4.2-22.2)	12.9 (7.2-22.3)	0.47	--	--

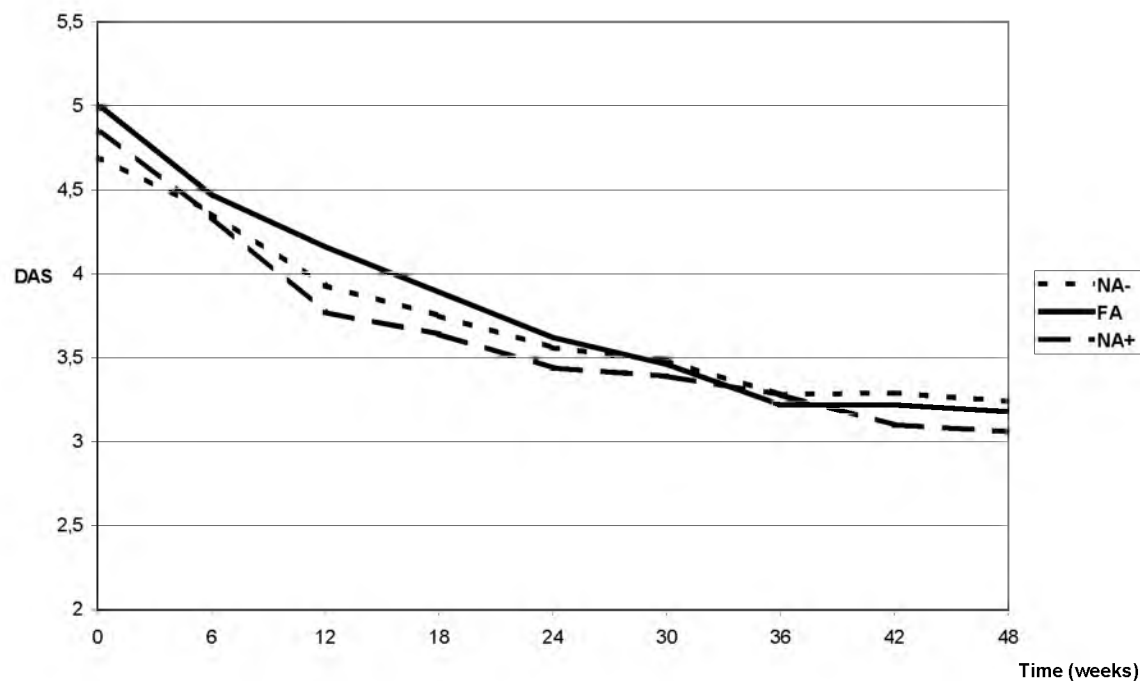
NA - : Patients with a net lower MTX dose than the guidelines proposed; NA + : Patients with a net higher MTX dose than the guidelines proposed; FA : Patients with full guideline adherence. DAS: Disease Activity Score. No. : Number. Values are counts (column percentage); mean (standard deviation); median (interquartile-range). * Good or moderate response.

Graph 1 Differences of cumulative MTX dose over time.

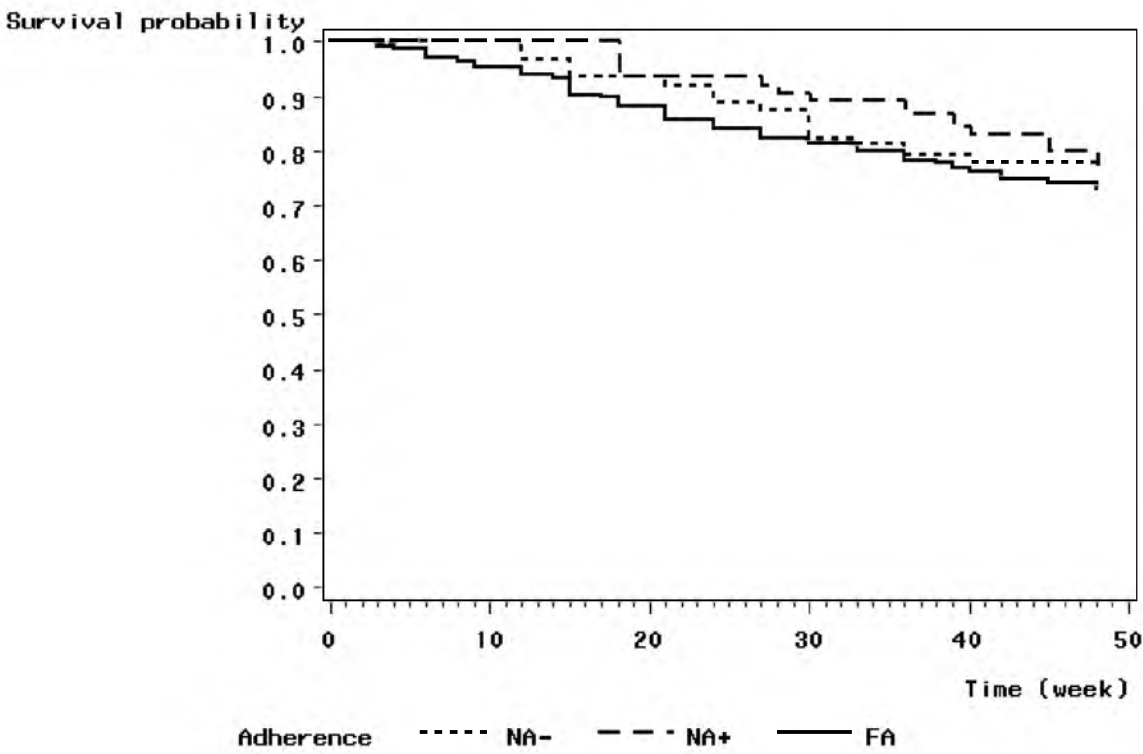


MTX cumulative dose differences are calculated as differences of the non-adherence groups from the full adherence group. Differences are presented as medians.

Graph 2 Course of disease activity (DAS) over time.



Graph 3 Survival analysis, dropout due to severe adverse events.



Chapter 8

Rheumatologists' opinion on the feasibility of a measurement-feedback system in RA – and the influence of motivation.

Jaap Fransen¹, Synove Daneel¹, Thomas Langenenegger² and Beat Michel¹
for the members of the Swiss Clinical Quality Management
in Rheumatoid Arthritis (SCQM).

¹ Department of Rheumatology and Institute of Physical Medicine,
University Hospital Zurich, Switzerland.

² Adelheid Page Rehabilitation Hospital, Unteraegeri, Switzerland.

Abstract

Objective To assess rheumatologists' opinion about feasibility of a measurement-feedback system in RA and to analyse if motivational aspects play a role in valuing and use.

Methods A survey sample (N=105) was randomly selected from participants of a measurement-feedback system. A survey questionnaire assessed opinions on system outcome, structures and processes, motivation and overall satisfaction. Survey results are given descriptively, motivation groups were compared.

Results Overall response was 62%. The system is generally perceived to fulfil its aims, but the effort to be made is rated less positive. Rheumatologists may have either "science/obligation" or "individual patient evaluation" as motivation. Latter rheumatologists are more satisfied with the measurement-feedback system, perceived its feasibility as better, and made more use of it.

Conclusion Motivation for participating in a measurement-feedback system has a significant impact on overall satisfaction and use. Influencing motivation and specific reduction of effort might increase overall acceptance.

Introduction

The Swiss Clinical Quality Management in RA (SCQM) is a co-operation of Rheumatology departments of all university- and most major hospitals, and rheumatologists in private practice throughout Switzerland. Since 1997, the SCQM provides rheumatologists with a measurement-feedback system for monitoring the disease course of individual RA patients. Using knowledge of the disease course, rheumatologists can optimise drug therapy to reduce disease activity and prevent progression of joint damage and concurring disability [1]. The accumulated cohort data can be used to monitor disease course and burden of disease, and to explore the use of different treatment strategies and predictors for outcome [1]. The long-term goal of the SCQM is to enable all RA patients to have the optimal treatment, thus minimising the burden of the disease to the patient, to his or her social environment and to society [1].

Use of the measurement-feedback system is free of charge. Rheumatologists mail standardised clinical and laboratory assessments, patient questionnaires, and annually X-ray files, to the SCQM co-ordination centre, where the data are processed and stored in a database [2]. A comprehensive paper feedback report on the individual patient is mailed back to the rheumatologist. It contains a graphical and numerical display of the course of disease activity (Disease activity score, Rheumatoid arthritis disease activity

index, joint counts, pain), disability (Health assessment questionnaire) and joint damage (Ratingen X-ray score), laboratory values (e.g. ESR, blood count, alanine aminotransferase) and an overview of past and current drug therapy; see [2] for more detail. A difference is made between annual assessments and intermediate assessments.

The rheumatologists are reminded to perform annual assessments by the co-ordination centre. Annual assessments are full assessments including disability and joint damage. Reduced intermediate assessments of disease activity and laboratory values are performed throughout the year at discretion of the rheumatologist.

For widespread use in clinical practice, a measurement-feedback system should be feasible and appreciated as support in daily clinical practice and upon making treatment decisions. For studying its efficacy, it is important that compliance to a measurement-feedback system is high, as it makes no sense to study the effectiveness of an intervention that cannot be adhered to. In an implementation study (submitted) as well as in daily practice it was noticed that some rheumatologists did not actively make use of the measurement-feedback system, while others intensively used it. To be able to improve broader use of the measurement-feedback system, opinion on feasibility was assessed among a random sample of rheumatologists with at least some experience with the system. In addition, it was analysed whether motivational aspects play a role in valuing and use of the measurement-feedback system.

Methods

There were 267 rheumatologists using the measurement-feedback system in 2001. A sample (N=105) was drawn with a random number generator, stratified for private practice (n=40), regional hospital (n=25), and university hospital (n=40). A numbered survey questionnaire was mailed to the sample members early 2002, with a pre-addressed and prepaid return envelope enclosed. After 4 weeks, a new questionnaire was sent to non-responders.

The survey questionnaire was developed to assess opinion on feasibility of the measurement-feedback system in 10-15 minutes. An intervention can be thought of as feasible, if it sufficiently meets three points: 1) The degree in which the intervention meets its stated aims (outcome); 2) The practicability of the structures needed to execute the intervention (structure); 3) The acceptability of the way the intervention is dealt with (process). These points are covering the three main aspects of clinical quality as formulated by Donabedian: outcome, structure and process [3]. There will be an, often

implicit, trade-off of perceived advantages and drawbacks, leading to an overall valuing which can be expressed as satisfaction with the intervention.

The survey questionnaire contained 23 feasibility items with a Likert-scale response format 1-6, where 1 is the positive and 6 the negative extreme [4]. Additionally included were questions on practice size, motivation to participate in the system, perceived advantages and drawbacks and overall satisfaction. The questionnaire was pre-tested for apprehension and completeness.

Actual use of the measurement-feedback system in 2001 was determined from the SCQM database for every responding rheumatologist, by the number of RA patients included in the system, and by the number of them having any intermediate visits.

The survey question about the motivation to participate in the measurement-feedback system was qualitatively assessed, and clearly two motivational groups could be formed. Wordings like “evaluation”, “long-term follow-up”, “decision support”, etc. were classified as describing factors directly related, or “internal”, to the patient-physician relationship. Those mentioning “obligatory”, “hospital custom”, “science”, or left the question blank, were classified as expressing factors not directly related, or “external”, to the patient-physician relationship. Internal and external do not refer to a specific psychological construct.

Likert scale item responses were analysed descriptively. Differences between both motivation groups were analysed using item response means and the 2-sample t-test, for ease of comparison. Difference in overall satisfaction was analysed using Fischer’s exact test. Differences in use were analysed using the 2-sample Wilcoxon test.

Results

Response

Of the 105 questionnaires sent, 55 were returned initially and 10 after the reminder. Overall response was 62%, with 65% for private practitioners, 80% in regional hospitals, and 47% in university hospitals.

Outcome

The responses to the survey items are described in the left-hand part of Table 1.

The majority of respondents (strongly) agreed with the statements, that the measurement-feedback is useful in practice, provides a better insight in the disease process, and is helpful in monitoring treatment effects and disease course.

The majority could not fully share the statement that the feedback on disease activity influences their treatment decisions. However, many respondents stated that feedback makes them feel more certain about treatment decisions taken.

Structure

Most respondents perceived the information as clearly presented and of relevance. But a large minority did not fully agree that the information is complete enough for most patients to base DMARD treatment decisions on.

The time necessary to perform the assessments of disease activity was generally perceived as acceptable. A majority of respondents could not confirm the statement that all in all time could be saved later, or that the time spent is in balance with time gained. This especially concerns the larger annual assessments, not the intermediate assessments of disease activity.

Process

A majority of the respondents stated that they always, or most of the times, showed and explained the feedback to the patients, and a majority had a look at it upon receipt. According to most respondents, they do not regularly assess disease activity between the annual assessments. A majority of respondents agreed that performing annual X-rays is acceptable for most cases. However, this was the subject of many drawbacks (see later). The number of respondents that rated not feeling positive about the effort made is considerably large.

Motivation

The rheumatologists' motivation, described in internal or external factors, makes a difference in the rating of overall satisfaction (Table 1). There is a tendency for rheumatologists with "internal" motivation, to rate feasibility issues more positive than rheumatologists with an "external" motivation (Table 1). E.g. the time spent for an intermediate visit was rated more positive by rheumatologists with "internal" motivation. Rheumatologists with "internal" motivation also show and explain the feedback report to their patients more often.

Both motivation groups rated the time consumption the same. The majority (66%) stated that they used additionally 1-10 minutes when assessing disease activity. For the larger annual visits, the rating of the extra time spent ranged from 5-10 minutes to >20 minutes. 46% stated that the feedback report was received between 10-18 days after the patient visit, 33% did not know when it arrived.

Use

The rheumatologists who described an “internal” motivation, included more of their RA patients, and performed intermediate assessments in the larger proportion of these patients (Table 2). The differences in use could not be explained by a small difference between centres, in the hospitals 50% described an “internal” motivation, in private practice 70%.

Advantages and drawbacks

Fifty respondents described advantages and drawbacks. The advantages most often mentioned were: long-term evaluation/follow-up disease course (n=30); decision making/treatment quality (n=25); clear and comprehensive overview (n=25); use of objectivable, systematic assessments (n=23); patient and physician communication aid (n=17). Drawbacks most often described were: time consumption (n=21); little or no consequences for decision support (n=11); inefficient paper bureaucracy (n=9); annual X-rays for all (n=8); problems with acquiring patient co-operation (n=7); lag-time of the feedback report (n=5); no scientific outcome (n=5); problems with the patient questionnaires (n=5); patients are a selection of the RA population (n=4); difficulties with the standardised medication documentation (n=4).

Discussion

With the survey a return rate of 62% was reached, which can be considered satisfactory for (anonymous) surveys. The low return rate in university hospitals may be caused by the large physician turnover in teaching hospitals.

In this sample, the measurement-feedback system was not frequently used, which is not in agreement with its purpose to serve as a decision-support to make regular adaptations of DMARD therapy. Rheumatologists that participate in the measurement-feedback system for the evaluation of their individual RA patients (“internal motivation”) perceived the measurement-feedback system as more useful, were more satisfied and less bothered by the time consumption, and also made more often use of it. On the contrary,

if the measurement-feedback was not perceived as useful in treating the own patients (“external motivation”), feasibility was judged in a more negative way, and use is less. Nevertheless, it is supportive that the SCQM measurement-feedback system was generally perceived to fulfil its aims, and this was clearly stated in the advantages. However, overall satisfaction was not optimal, 70% (“internal”) and 40% (“external”), respectively. A satisfaction rate of 80% is not unusual for many situations in health care [5]. Thus, there is reason to believe that satisfaction should be improved, ideally approaching 100%, in all fields of outcome, structure and process. However it can be noticed from Table 1 that dissatisfaction concentrates on time costs and decision making.

The major advantage of instituting a standardised measurement-feedback system in general is that important aspects of disease can be expressed in a more objective way. Then, treatment decisions can be based on (semi-) objective data, it is easier to set and check treatment targets, treatment results can be compared with trial results, and the data may serve as communication aid [6]. However, it should be clear that measurement-feedback is meant to be a support for decision making, it adds to, but does not replace, clinical thinking and other relevant sources of information. This principle should be clearly communicated to users. Major drawbacks of a standardised measurement-feedback system are, that it may be felt as a reduction of freedom which makes the effort feel as burden, it may make people think that numbers are considered more important than patients, and rheumatologists may be afraid for being controlled and judged on some abstract treatment result. Further, measurement-feedback systems may be quite costly; e.g. an assessment costs the SCQM between 25 and 65 Euro. Even if rheumatologists do not pay for the SCQM measurement-feedback system, time means also money for a lot of physicians.

The SCQM measurement-feedback system was initially developed in a consensus process with key-players among Swiss rheumatologists. From surveying current users, several practical points can be derived to be able to increase use of a measurement-feedback system for RA. Concerning outcome: Regular workshops on performance of assessments and interpretation of scores, and how these can be used for decision making, could be provided. Especially for non-laboratory assessments, it can be important to raise familiarity, agreement, self-efficacy, and positive outcome expectancy [7]. Concerning structure: Easy access to treatment guidelines and a possibility for consultancy in difficult cases could be established, apart from the provision of feedback alone. Electronic solutions could replace paper ones to reduce effort and to increase

speed. Patient questionnaires may be collected by a co-ordination centre, instead of the rheumatologists. Concerning process: The performance of X-rays could be changed, e.g. less frequent in established disease, and more frequent in early disease. Immediate feedback, instead of a time lag of several days, is a great advantage for decision making. Then, treatment decisions can be made in presence of the patient, which may enhance patient compliance. An easy to understand written and illustrated patient information may very well lighten the unpleasant task of persuading patients to join the measurement-feedback system. In our opinion, it is of major importance that the aim of a measurement-feedback system should clearly be communicated, and that its use for evaluating individual RA patients should be stimulated. For the commitment to science, it seems important to create "data ownership" and to make regular short communications for all data contributors, e.g. in a newsletter.

Acknowledgements

We are grateful to all rheumatologists that found time to response to the survey.

We wish to thank Leanne Pobjoy for her assistance in the preparation of the manuscript.

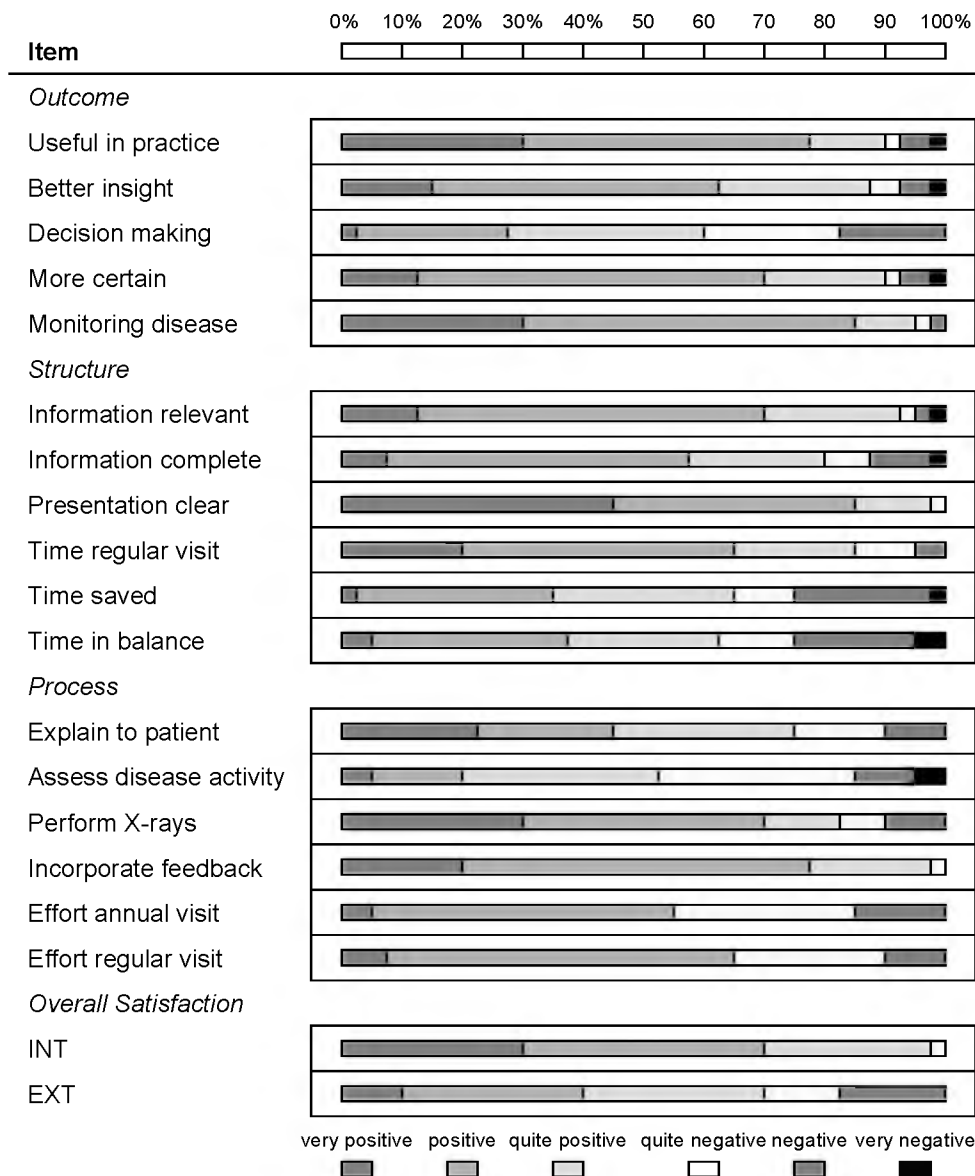
This work is part from a project sponsored by a grant from the Swiss Health Authorities (BAG).

The members of the Swiss Clinical Quality Management in Rheumatoid Arthritis are: Dr. I. Büchler, Dr. B. Christen, Dr. A-M. Chamot, G. Dalvit, Dr. A. Forster, PD Dr. P.A. Guerne, Dr. T. Langenegger, Prof. Dr. BA. Michel, Dr. L. Schmid, Dr. H. Schwarz, Prof. Dr. M. Seitz, Prof. Dr. A. So, PD Dr. R. Theiler, Prof. Dr. A. Tyndall, PD Dr. D. van Linthoudt, Prof. Dr. P. Villiger

References

1. Stucki G, Sangha O. Clinical quality management: putting the pieces together. *Arthritis Care Res* 1996;9(5):405-12.
2. Uitz E, Fransen J, Langenegger T, Stucki G. Clinical quality management in rheumatoid arthritis: putting theory into practice. *Swiss Clinical Quality Management in Rheumatoid Arthritis. Rheumatology* 2000;39(5):542-9.
3. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Quart* 1966;44:166-203.
4. Streiner DL, Norman GR. *Health Measurement Scales. A practical guide to their development and use.* Oxford: Oxford University Press, 1999.
5. Di Palo MT. Rating satisfaction research: Is it poor, fair, good, very good or excellent? *Arthritis Care Res* 1997;10(6):422-30.
6. Fransen J, Stucki G, Van Riel PLCM. The merits of monitoring: Should we follow all our RA patients in daily clinical practice? *Rheumatology* 2002;41:601-4.
7. Cabana MD, Rushton JL, Rush AJ. Implementing practice guidelines for depression: Applying a new framework to an old problem. *Gen Hosp Psychiatry* 2002;24(1):35-42.

Table 1 Survey item responses.



INT: evaluation as motivation; EXT: obligation or science as motivation

Total	INT	EXT	P-value
2.1	1.7	2.7	0.001
2.4	2.0	3.1	<0.0001
3.2	2.8	3.8	<0.0001
2.4	2.0	2.8	0.008
2.0	1.7	2.3	0.02
2.3	2.1	2.5	0.04
2.8	2.4	3.2	0.02
1.7	1.7	1.7	0.88
2.4	2.0	2.9	0.005
3.2	3.1	3.4	0.44
3.2	2.9	3.7	0.02
2.7	2.5	2.9	0.26
3.4	3.5	3.3	0.57
2.3	2.1	2.5	0.37
2.0	1.8	2.4	<0.0001
2.6	2.3	2.8	0.02
2.3	2.2	2.5	0.14
			0.006

Table 2 Use of the measurement-feedback system in 2001.

	INT	EXT	p-value
n=	38	28	
% included	75%	11%	0.018
% followed	37%	0%	0.027

Median percentages. P-values from 2-sample Wilcoxon test.
 INT= Motivation by internal factors, EXT = by external factors.

Chapter 9

Design and analysis of a RCT on the effects of clinical decision support on physician performance and health outcome in the management of RA.

Jaap Fransen¹, Jos Twisk², Marjonne Creemers¹ and Piet van Riel¹.

¹ Department of Rheumatology, University Medical Center Nijmegen, The Netherlands.

² Institute of Research in Extramural Medicine, Faculty of Medicine, Free University Amsterdam, The Netherlands.

Abstract

Objective A proposal is made for design and analysis of a randomized controlled trial to evaluate the effects of a clinical decision support system on physician performance and health outcome in RA.

Methods Sample size calculations for classical and cluster RCT designs were performed using a dichotomous and a continuous outcome measure, based on cohort data.

Results A cluster RCT comparing decision support with usual care is the most appropriate design. The proportion of patients with rheumatoid inflammation under control is the preferred primary outcome measure, for which the Disease Activity Score can be used. A sample size of 268 RA patients would be needed to detect a between-group difference of 30%. The required sample size is 2.5 times larger than when clustering is ignored. The data can be analyzed using multi-level analysis.

Conclusion As cluster RCTs are increasingly being used, awareness should be raised of their design and analysis. The methodological considerations in this paper are applicable to similar research objectives in Rheumatology.

Introduction

For the management of RA, there is general agreement that rheumatoid inflammation should be controlled as soon as possible, as completely as possible and for as long as possible, consistent with patient safety [1,2]. When accepting that the goal of RA treatment is to reach optimal control of rheumatoid inflammation, it is clear that rheumatoid inflammation should continuously be evaluated. Then, the treatment program can be adjusted from both perspectives of benefit and harm [3]. The combination of systematic evaluation and clinical guidelines could be a valuable decision support in optimizing the management of RA [4]. The effects of such decision support should preferably be studied using a randomized controlled trial (RCT) design. In this paper, a proposal is made for design and analysis of a RCT to evaluate the effects of a clinical decision support system on physician performance and health outcome in the management of RA. The design of a cluster RCT, the choice of the relevant outcome and outcome measure, an approach for statistical analysis, and sample size calculations are emphasized. Solutions to prevent bias in this particular design and ethical considerations are described.

Choice of the trial design

In a classical RCT design, patients from single physicians are randomized to receive either an experimental or a control intervention, while patients and physicians are blind for treatment allocation. A cluster RCT design is appropriate when interventions, like guidelines or decision support, are primarily directed to the physicians instead of the patients. In that case, the physicians cannot be blinded. To avoid contamination, it is the physicians or practices that are randomized, and all their study patients (the cluster) will receive the same intervention [5]. Then, it is not appropriate anymore to regard patients as independent. The degree of dependency of patients in clusters is indicated by the height of the intraclass correlation coefficient (ICC), an ICC equal to zero indicates independency. The higher the ICC, the less 'unique' information is contributed by the single patient and the more the power of the study is reduced. Therefore, it is necessary to correct for the effect of clustering in sample size calculations and analysis [6,7].

Choice of the relevant outcome

Decision support, such as guidelines, aims to persuade physicians to change their practice behavior [5]. Therefore, when studying decision support it is physician performance that is the outcome of interest [5]. Concerning the management of RA, a measure of physician performance would thus reflect the adequacy of decisions on medication. However, it is difficult to use treatment decisions as an outcome measure, because in the management of RA visit frequency is normally not standardized and a large number of treatment options are available. As guidelines may alter visit frequency, more opportunity for guideline (non-) adherence may be present in one of the trial arms, possibly leading to bias. Further, with many treatment options available, judgment of the adequacy of treatment decisions is not straightforward, and single visits may not be judged in isolation. Since the aim of RA management is to control rheumatoid inflammation [1], the level of rheumatoid inflammation or the proportion of patients with adequately controlled rheumatoid inflammation could alternatively be used as primary outcome. However, it must be noted that these are essentially measures of health outcome, and are no direct measures of physician performance. In order to explain possible changes in outcome, physician performance should be documented. Secondary outcomes that can be considered are toxicity, disability, joint damage, quality of life, satisfaction with care, resource use and direct costs.

Choice of the primary outcome measure

Multiple measures for the clinical assessment of rheumatoid inflammation are available, but all are approximating rather than measuring the underlying disease process. The disease activity score (DAS28) is a well validated measure, calculated from the results of ESR, a swollen joint count, a tender joint count and a general health item [8]. The DAS28 provides a single index reflecting the level of disease activity. It is a continuous measure having a Gaussian distribution, which is an advantage for analysis and power calculations. The proportion of patients with adequately controlled disease activity can also be derived, e.g. by using a dichotomy with a cut-off point of $\text{DAS28} \leq 3.2$ [8], but power may be lost by dichotomizing a continuous measure.

Statistical Analysis

The proportion of patients with rheumatoid inflammation under control is a simple combination of patients from one rheumatologist into a single summary statistic [5]. Then, a two-sample t-test with or without weighting for cluster size [5,9], an adjusted chi-square test or computation of an odds ratio with adjusted confidence intervals [5] can be carried out. Similarly, the mean change in DAS28 per practice can be used as summary statistic, and the two-sample t-test, with or without weighting for cluster size, can be carried out [5,9]. The rheumatologist, or practice, was the unit of randomization and is also the unit of analysis in these techniques, with the advantage of being relatively simple. Alternatively, for dichotomous (low or high DAS28) as well as for continuous (change in DAS28) outcomes, multi-level analysis (MLA) [9] can be carried out. MLA is not performed on cluster level (rheumatologist or practice) as above, but on individual patient level while correcting for the dependency within clusters. MLA is quite a complex technique. However, its' advantage is, as with multiple regression, that it is much easier to correct for differences in comparability between trial arms and that e.g. a correction for the level of DAS28 at baseline can be made [9].

Sample size when the outcome is dichotomous

Ignoring the clustering of the data, the sample size per trial arm was calculated with the usual level for $\alpha = 0.05$, and a power $(1 - \beta)$ of 0.90, using $n = 2(Z_{\alpha/2} + Z_{\beta})^2 p(1 - p) / (p_1 - p_0)^2$; see [10] for an introduction. $Z_{\alpha/2}$ is the z-value on the standard normal distribution corresponding with the chance α to find a difference when in truth none exists; if $\alpha=0.05$ (two-sided) then $Z_{\alpha/2} = 1.96$. Z_{β} is the z-value on the standard normal distribution corresponding with the chance β of not finding a difference, when in truth it exists; if

$\beta=0.10$ then $Z_\beta=1.28$. The proportion of patients with disease activity under control at the end of the trial in experimental group and control group are denoted as p_1 and p_0 , while p is the pooled proportion of patients with disease activity under control $(p_1 + p_0)/2$.

A sample resembling the target population was taken from an existing cohort [11], and consisted of 570 RA patients and 50 rheumatologists (a mean of 11.4 RA patients per rheumatologist). The proportion of patients with a $\text{DAS28} \leq 3.2$ was 20%. It was hypothesized that it was a manageable target to reach low disease activity ($\text{DAS28} \leq 3.2$) in 50% of the RA patients in the experimental group, p_1 was then defined to be 0.50 and p_0 as 0.20. Accordingly, the number of patients needed per trial arm was calculated as $n=53$.

It was already indicated that the clustered nature of data reduces power. Therefore, sample size has to be increased with a factor known as the design effect (DE). The DE is the ratio of sample size with, and sample size without adjustment for clustering, and can be calculated using $\text{DE} = 1+(m-1)p$, where m is the average cluster size and p denotes the ICC. The ICC was calculated using the data from our cohort sample [11]. An ICC of $p = 0.13$ was obtained, calculated as $p = \sigma_b^2 / \sigma_b^2 + (m - 1)\sigma_w^2$, where σ_b^2 is the between cluster variance, σ_w^2 is the within cluster variance, and m represents the average cluster size [9]. The resulting design effect is $\text{DE} = 1+(11.4 - 1) \times 0.13 = 2.35$. The sample size needed follows from multiplying the unadjusted sample size with the DE: $53 \times 2.35 = 125$ patients per treatment arm. For our cohort sample this would mean that $(125 / 11.4)$ 11 rheumatologists would be needed. Calculation of DE assumes equal cluster size. However, usually cluster sizes (practice sizes) are unequal. Then, again power is lost and the DE calculated is too small [7]. A correction should be applied, for which the use of minimum variance weights is recommended [7], using

$$\text{DE} = mM / \sum_{i=1}^M m_i / 1 + (m_i - 1)p,$$

where m is the average cluster size, M is the number of clusters, and m_i is the observed cluster size. This increases the design effect to $\text{DE} = 2.53$, leading to a sample size per treatment arm of 134 patients requiring 12 rheumatologists.

Sample size when the outcome is continuous

Again, sample size needed per trial arm was first calculated while ignoring the effect of clustering, with $\alpha = 0.05$ and $1 - \beta = 0.90$ using $n = 2(Z_{\alpha/2} + Z_\beta)^2 \sigma^2 / (m_1 - m_0)^2$. The standard deviation for the DAS28 ($\sigma=1.48$) was again estimated from our cohort sample

[11]. A relevant difference between mean group changes ($m_1 - m_0$) of 1.2 in the DAS28 was chosen [8]. As the mean DAS28 in the cohort sample was 4.4, it can be expected that with a mean improvement of 1.2 in the intervention group, approximately 50% of those patients will have low disease activity ($\text{DAS28} \leq 3.2$) [8]. Accordingly, the number of patients per treatment arm was calculated as $n=32$. The ICC was again calculated from our cohort sample, according to $p = \sigma^2_b / (\sigma^2_b + \sigma^2_w)$ [12], resulting in an $\text{ICC}=0.25$. The design effect was calculated as $\text{DE} = 1 + (11.4 - 1)0.25 = 3.60$. The resulting sample size per trial arm is $32 \times 3.6 = 115$ patients, requiring 11 rheumatologists. The design effect adjusted for unequal cluster size was calculated as $\text{DE} = 3.79$ [7]. The sample size has to be increased to 121 patients per trial arm, needing 11 rheumatologists.

Choice of the interventions

An experimental intervention could be a computerized clinical decision support system (CDSS), allowing for the systematic evaluation of rheumatoid inflammation and the provision of clinical guidelines. A threat for the validity of the study is non-use of the CDSS by the rheumatologists, which can especially occur when rheumatologists treat only a small number of RA patients. Adherence can be enhanced by raising familiarity and agreement [13]. The CDSS must be practical, include feasible measurements, and produce immediate and meaningful results. It is important to include a run-in period for CDSS use and training in the study.

A control intervention ideally should resemble 'usual care'. However, it can be difficult to control for contamination, as it can easily occur by information in literature, by exchange between rheumatologists, or by informed consent procedure.

Recruitment of rheumatologists and patients

The study intervention requires a high degree of participation and willingness to change practice style for rheumatologists. The influence of losing clusters, e.g. by recruitment failure or by drop out, on power can be large if the number of clusters is small (e.g. 20) [6]. Thus an appropriate recruitment strategy and adequate selection criteria for rheumatologists as well as ongoing motivation by the study management are needed. Because rheumatologists will play an important role in patient recruitment, it is important to use clear patient selection criteria. If randomization is stratified using patient variables (e.g. level of disease activity and disease duration), patient recruitment has to take place before randomization. That may also be helpful in the prevention of selection bias,

because rheumatologists randomized to usual care might be less motivated to recruit patients, or might recruit patients with a different profile [14].

Blinding

With complex interventions, such as decision support, that are directed to the physicians instead of the patients, it is impossible to keep physicians blind. It is also difficult to keep patients blind, but an effort must be made to keep patients naive towards outcome expectancy. In concordance with clinical practice, the assessments of disease activity in the CDSS trial arm, like the DAS28, can be performed by the unblind rheumatologist. For statistical analysis of this kind of RCT however, only assessments can be used that are made by an independent blinded assessor, e.g. a nurse practitioner. The assessment time frame should be identical for both trial arms.

Follow-up

Disease Modifying Antirheumatic Drugs (DMARDs) are the therapy of choice in RA. For the evaluation of the efficacy of DMARDs, a follow-up duration of 12 months is generally seen as adequate. Clear reductions in inflammatory activity can be expected within 3-6 months with the DMARDs now available [2]. The primary outcome of CDSS can therefore be analyzed at 6 or 12 months. When the course of disease activity over time is also of interest, it is advisable to perform blinded outcome assessments at least every 3 months.

Ethical considerations

In the proposed cluster RCT, treatment options are randomized and patients are individually treated and followed. Then, it is necessary to obtain informed consent by the individual patients [15]. Only in the special case that an intervention cannot be targeted at an individual but only at the cluster as a whole (e.g. special medical education) while outcome is on practice level (e.g. number of adequate referrals), consent may be obtained only from the person responsible for the cluster's well being (e.g. a rheumatologist) [15].

Discussion

A cluster RCT is the most appropriate design for studying the effects of CDSS in the management of RA. Using the proportion of patients with a $\text{DAS28} \leq 3.2$ as outcome, it was estimated that a sample of 268 RA patients and 24 rheumatologists would be

needed, which is 2.5 times larger than when the clustering effect is ignored. Multilevel Analysis (MLA) may be a useful approach for statistical analysis.

Generally when studying decision support, such as guidelines, it is not necessary to study the effect on health [5]. If included guidelines are based on sound evidence, it is already known that targeted behavior will be beneficial [5]. As an example, 'efficacy' of guidelines to improve folate supplementation in addition to MTX can be evaluated by simply counting the number of correct prescriptions. In case of CDSS in RA, physician performance is more difficult to measure. Therefore, the proportion of patients with low disease activity ($\text{DAS28} \leq 3.2$) was proposed to approximate physician performance. However, as there is no 'gold standard' to measure RA disease activity, other outcomes can be considered, e.g. the proportion of patients in remission, the proportion of responders, change in disease activity, or time-integrated disease activity. Further, depending on the target population, other relevant differences and distributions may be chosen in the formulas for power calculation, leading to other sample size estimations. Also, the magnitude of the ICC may be different for other populations and measures. Most of the times power is lost when dichotomizing a continuous measure, and the estimated sample size will therefore increase. In our example, no large difference in sample size between use of the DAS28 as a continuous measure or as a dichotomy appeared. However, when initially clustering was ignored, the required sample sizes ($n=32$ versus $n=53$) showed that indeed power was lost when dichotomizing the continuous DAS28. The estimated sample sizes became more similar when corrected for the design effect, due to differences in the height of the observed ICC's.

In the past, consequences of clustered data have largely been ignored in clinical studies, and as a result, many published studies are underpowered [16]. Currently, cluster RCTs are increasingly being used, especially in public health- and general practice research. Also in Rheumatology, practitioners and researchers should be aware of the consequences of clustering for trial design and analysis. The methodological considerations described in this paper are applicable to similar research objectives in Rheumatology.

References

1. Moreland LW, Russell AS, Paulus HE. Management of rheumatoid arthritis: the historical context. *J Rheumatol* 2001;28(6):1431-52.
2. Wolfe F, Cush JJ, O'Dell JR, Kavanaugh A, Kremer JM, Lane NE et al. Consensus recommendations for the assessment and treatment of rheumatoid arthritis. *J Rheumatol* 2001;28(6):1423-30.
3. Guidelines for the management of rheumatoid arthritis. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. *Arthritis Rheum* 1996;39(5):713-22.
4. Van Gestel AM, Van Riel PLCM. Evaluation of early rheumatoid arthritis disease activity and outcome. *Baillieres Clin Rheumatol* 1997;11(1):49-63.
5. Bland JM. Sample size in guidelines trials. *Fam Pract* 2000;17(Suppl1):S17-20.
6. Yudkin PL, Moher M. Putting theory into practice: a cluster randomized trial with a small number of clusters. *Stat Med* 2001;20(3):341-9.
7. Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Stat Med* 2001;20(3):377-90.
8. Van Gestel AM, Prevoo MLL, Van 't Hof MA, Van Rijswijk MH, Van de Putte LBA, Van Riel PLCM. Development and validation of the European league against rheumatism response criteria for rheumatoid arthritis. *Arthritis Rheum* 1996;39(11):34-40.
9. Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol* 1994;140(3):279-89.
10. Du V Florey C. Sample size for beginners. *BMJ* 1993;306(6886):1181-4.
11. Fransen J, Uebelhart D, Stucki G, Langenegger T, Seitz M, Michel BA. The ICIDH-2 as a framework for the assessment of functioning and disability in rheumatoid arthritis. *Ann Rheum Dis* 2002;61:225-31.
12. Campbell MK, Mollison J, Grimshaw JM. Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med* 2001;20(3):391-9.
13. Cabana MD, Rushton JL, Rush AJ. Implementing practice guidelines for depression: Applying a new framework to an old problem. *Gen Hosp Psychiatry* 2002;24(1):35-42.
14. Moore H, Summerbell C, Vail A, Greenwood DC, Adamson AJ. The design features and practicalities of conducting a pragmatic cluster randomized trial of obesity management in primary care. *Stat Med* 2001;20(3):331-40.
15. Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomized controlled trial. *BMJ* 1999;318(7195):1407-9.
16. Campbell MK, Grimshaw JM. Cluster randomized trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. *BMJ* 1998;317(7167):1171-2.

Chapter 10

Summary

Introduction

The last decade, important developments in the management of rheumatoid arthritis (RA) occurred. It was recognised that RA should be treated as early as possible with DMARDs, more effective treatment options became available, and progress was made in the assessment of treatment effects.

To support the management of RA, the Swiss Clinical Quality Management in Rheumatoid Arthritis (SCQM) was established. Core activity of the SCQM is to provide rheumatologists with a measurement-feedback system to monitor the course of RA. Using the feedback, drug therapy can be optimised to reduce disease activity, leading to prevention of joint damage and concurring disability.

The main subject of this thesis is to study if measurement-feedback, or decision support, is effective in the management of RA.

Chapter 2

There is general agreement that rheumatoid inflammation should be controlled as soon as possible, as completely as possible, and that control should be maintained for as long as possible, consistent with patient safety.

Accepting that the goal of treatment is to reach optimal control of rheumatoid inflammation, it is clear that rheumatoid inflammation should regularly be evaluated. However, regular and systematic monitoring of rheumatoid inflammation is not commonly used in daily clinical practice. Beneath practical barriers, the most important barrier is: how does a clinician know when rheumatoid inflammation is optimally controlled or in remission? What are the criteria to state that inflammatory activity is under control?

There is no gold standard to assess the severity of rheumatoid inflammation, thus it has to be estimated. The disease activity score (DAS) and the EULAR response criteria are suited to aid in determining and evaluating actual status and change in status, particularly when applied to individual RA patients. The reason is, that the EULAR response status depends on the absolute level of DAS reached. A low level of DAS over time reduces the probability of progression of radiological visible joint damage. How the DAS can be used for dose titration is demonstrated in 2 studies in RA, using a step-up and a step-down protocol.

It is concluded that, to evaluate if treatment goals are reached and to support accomodation of the treatment program, the management of RA patients in daily clinical practice should include systematic and regular evaluation of rheumatoid inflammation.

For determining status and change of individual RA patients in daily clinical practice, core set measures and response criteria may be used.

Chapter 3

Clinical Quality Management (CQM) in RA aims at reduction of inflammatory activity and pain in the short-term, and damage, and consequent disability, in the long-term.

Rheumatologists are provided with a measurement feedback system with which they can regularly follow their patients. Inflammatory activity is measured with the Disease Activity Score (DAS28) and the Rheumatoid Arthritis Disease Activity Index Questionnaire (RADAI), damage with an X-ray score, disability with the Stanford Health Assessment Questionnaire (HAQ). Feedback is used to optimise therapy, which in the short term allows the activity of the inflammatory process to be adjusted or "titrated". In the long term, the therapy result for the individual patient is monitored by the course of disability and damage.

In this chapter, a series of cases is presented to illustrate the usefulness of the CQM system in the management of individual RA patients. In Case 1, it is illustrated how periods of increased inflammatory activity (flares) present in numerical values and in graphs, and how the principle of titration of inflammatory activity works in practice.

Case 2 illustrates that while in mild to moderate RA titration of disease activity towards a DAS28 below 3.2 may be achieved using a single DMARD, patients with more severe RA may require combination therapy. The next 2 cases illustrate some regularly encountered situations requiring individualised interpretation of measures, reminding us that treatment of RA is still an art rather than a cookbook approach.

It is concluded that CQM in RA may be helpful when making decisions about adjustment of treatment, and to document and communicate these decisions based on quantitative data.

Chapter 4

The goal of the Rheumatoid Arthritis Disease Activity Index (RADAI) is to provide an easy to use assessment of disease activity. It is a self-administered questionnaire that combines five items into a single index: current and past global disease activity, pain, morning stiffness and a joint count. In an earlier study, the single index approach of the RADAI was found to be valid, basing on the high association of the RADAI with clinically assessed joint synovitis and the acute-phase response, the high internal consistency,

and the loading of the RADAI items on a single factor. However, the sample of RA patients used was relatively small (N=55).

A cross-sectional sample of 584 rheumatoid arthritis (RA) outpatients was used to assess the internal consistency and the convergent validity of the RADAI. The sample represented a wide range of disease duration and levels of disease activity and disability. Cronbach's α was 0.87, supporting the summation of the items into a single index. The index correlates best with physicians' global assessment of disease activity ($r=0.59$; $p<0.0001$), HAQ ($r=0.55$; $p<0.0001$) and the tender joint count ($r=0.55$; $p<0.0001$). Correlation with ESR is low ($r=0.27$; $p<0.0001$). The RADAI and DAS28 are correlated ($r=0.53$; $p<0.0001$), but agreement is low.

It is concluded that the RADAI is valid to assess disease activity in RA patients. However, the RADAI may not automatically replace other measures of disease activity, such as the DAS28.

Chapter 5

In this study, the responsiveness of the Rheumatoid Arthritis Disease Activity Index (RADAI) is assessed for increases in disease activity in rheumatoid arthritis (RA) patients, with the occurrence of a flare of disease activity as an external standard.

Post-hoc analysis of data from a randomised double blind controlled trial of MTX versus Collagen II (N=92) was used. Responsiveness is analysed by the correlation of change in the RADAI with change in the Disease Activity Score (DAS28), and the ability of the RADAI to detect a flare by plotting a receiver-operating characteristic (ROC) curve and by the standardized effect size (SES). The contribution of the single RADAI items to the total change is analysed by absolute values, the standardized response mean (SRM), and correlation of item score change with the total RADAI score change.

The changes in the RADAI correlated highly ($r=0.70$, $p<0.0001$) with changes in the DAS28. The area under the ROC curve was 0.87 (95%CI: 0.78-0.95) for the RADAI, which was similar to the DAS28. The SES for the RADAI was 1.56, which was again similar to the DAS28. The RADAI items on "past global disease activity" and "morning stiffness" contributed the least to the total score change.

This study provides evidence that the RADAI is sensitive to detect relevant increases in disease activity in patients with RA. The RADAI may complement clinical measures in clinical studies, or may be used as a proxy for disease activity in epidemiological studies.

Chapter 6

With the help of a measurement-feedback system, the treatment strategy for individual RA patients can be adjusted to achieve optimal control of disease activity. The main objective was to study whether a measurement-feedback system is effective in reducing disease activity in RA patients.

Forty-eight rheumatologists and 264 patients participated in a controlled clinical trial. A 3-months control period was followed by a 12-months period, where feedback on disease activity, disability and damage was provided to the rheumatologist. Primary outcome measure was the Rheumatoid Arthritis Disease Activity Index (RADAI). The data were analysed using generalised estimating equations.

For 62% of the patients, the feedback system was used. DMARD changes occurred in 41% of the patients. In patients with high disease activity and feedback use ($n=70$), the RADAI decreased in the feedback period with -0.27 point per 30 days ($p<0.05$), as compared to the control period. Patients for which the feedback system was used had a better outcome than non-users.

Much more training regarding the use of a feedback system and outcome measures, as well as the inclusion of explicit treatment guidelines will be necessary to increase the clinical use of measurement-feedback and to possibly reduce disease activity for a larger number of RA patients.

Chapter 7

The objective was to study the influence of rheumatologists' guideline adherence on efficacy and toxicity in a 48 week, randomized controlled trial of methotrexate (MTX) with supplementation of folates or placebo in rheumatoid arthritis.

To reach an optimal dose of MTX, guidelines based on the disease activity score (DAS) and the occurrence of adverse events were applied. MTX was started with 7.5 mg/week and was raised every 6 weeks by 2.5 mg/week, until a good response was reached or adverse events occurred. Adherence to the guidelines was judged after the study had been completed. The course of the DAS was analyzed using generalized estimating equations; the occurrence of adverse events was analyzed with survival analysis corrected for time dependency.

In 51% of the 411 study patients the guidelines were always followed. In 25% of patients, non-adherence lead to lower doses and in 24% to higher doses of MTX than the guidelines had proposed. In the adherence group, the reduction of the DAS (mean 0.4; $p=0.0085$) was significantly larger compared with the 'low dose' group, the 'high dose'

and adherence groups did not differ (mean -0.07; $p=0.64$). The dropout due to severe adverse events appeared not to be different between the three groups.

The effect of prescribing MTX doses higher than the guidelines proposed seemed not to be more beneficial than guideline adherence, whereas the prescription of MTX doses lower than the guidelines proposed reduced efficacy without clear beneficial effect on toxicity. Guidelines assistive in determining the optimal efficacious MTX or DMARD dose may be useful in daily clinical practice.

Chapter 8

The objective of this survey was to assess rheumatologists' opinion about the feasibility of a measurement-feedback system in RA, and to analyse if motivational aspects play a role in perception and use.

A survey sample ($N=105$) was randomly selected from participants of a measurement-feedback system, stratified for private practice, regional hospital, and university hospital. The survey questionnaire assessed opinions on system outcome, structures and processes, motivation and overall satisfaction. In addition, it was determined for how many RA patients the measurement-feedback system was used last year. Survey results are given descriptively, the motivation groups were compared.

The overall response was 62%. The measurement-feedback system is generally perceived to fulfil its aims, but the effort to be made is rated less positive. Rheumatologists may have either "science/obligation" or "individual patient evaluation" as motivation. The latter rheumatologists are more satisfied with the measurement-feedback system, perceived its feasibility better, and made more use of it. Important practice points for improvement were derived from the users.

It is concluded that motivation for joining a measurement-feedback system has a significant impact on overall satisfaction and use. Influencing motivation and specific reduction of effort might increase overall acceptance.

Chapter 9

The combination of systematic evaluation and clinical guidelines aiming to control rheumatoid inflammation could be a valuable decision support in the management of RA. A proposal is made for design and analysis of a randomized controlled trial to evaluate the effects of a clinical decision support system on physician performance and health outcome in RA.

Sample size calculations for classical and cluster RCT designs were performed using a dichotomous and a continuous outcome measure. Population estimators were based on cohort data.

A cluster RCT comparing decision support with usual care is regarded the most appropriate design. The proportion of patients with rheumatoid inflammation under control is the preferred primary outcome measure, for which the Disease Activity Score (DAS28) can be used. A sample size of 268 RA patients and 24 rheumatologists would be needed to detect a between-group difference of 30%. The required sample size is 2.5 times larger than when clustering is ignored. The data can be analyzed using multi-level analysis.

As cluster RCTs are increasingly being used, awareness should be raised of their design and analysis. The methodological considerations for the example in this paper are applicable to similar research objectives in rheumatology.

Chapter 11

Samenvatting

Samenvatting

Reumatoïde artritis (RA) wordt met name gekenmerkt door chronische gewrichtsontstekingen, vooral aan beide handen en voeten. Ook andere gewrichten kunnen bij de ziekte betrokken raken. De ontstoken gewrichten zijn pijnlijk, gezwollen en soms stijf en daardoor lastig te gebruiken. Al in de eerste jaren na het begin van de ziekte kunnen de gewrichten onherstelbaar beschadigd raken, waarschijnlijk als gevolg van de voortgaande gewrichtsontsteking en een woekering van het gewrichtskapsel. Bij veel RA patiënten neemt de gewrichtsbeschadiging in de loop van de tijd toe. De beschadigde gewrichten kunnen gaan vervormen en vergroeien, veelal met pijn en problemen in het dagelijks functioneren als gevolg.

De precieze oorzaak van RA is nog onbekend. Het ontstekingsproces wordt vooral behandeld met bepaalde anti-reumatische medicijnen, met de familienaam "Disease Modifying Antirheumatic Drugs". Het doel van de behandeling met deze anti-reumatica is om het ontstekingsproces zo ver als mogelijk te onderdrukken en het voortgaande proces van gewrichtsbeschadiging te vertragen of zelfs te stoppen.

In plaats van behandeling met een enkel anti-reumaticum, worden voor een betere werking ook wel verschillende anti-reumatica in combinatie gebruikt. Ook worden vaak andere ontstekingsremmers en medicijnen om bijwerkingen te helpen voorkomen toegevoegd aan een behandeling met anti-reumatica.

Omdat veel RA patiënten langdurig meerdere medicijnen moeten gebruiken, heeft de reumatoloog een belangrijke taak in het instellen van de medicatie. Door regelmatig de activiteit van het ontstekingsproces te meten kan de reumatoloog bepalen of de medicamenteuze dosis verhoogd zou moeten worden, of dat de dosis weer omlaag kan. Bij het instellen van de medicatie speelt het ook een rol dat geprobeerd moet worden om geen, of in ieder geval zo min mogelijk, bijwerkingen te veroorzaken.

Het behandelingseffect kan dus op de korte termijn beoordeeld worden aan de hand van de ontstekingsactiviteit. Het behandelingseffect op de langere termijn kan beoordeeld worden met behulp van het dagelijks functioneren (middels een speciale patiënten vragenlijst) en de eventueel ontstane gewrichtsschade (middels röntgenfoto's van handen en voeten). Het systematisch beoordelen van ontstekingsactiviteit, functioneren en gewrichtsschade in de behandeling van RA is de laatste jaren meer in zwang geraakt, mede vanwege de komst van nieuwe, sneller en beter werkende anti-reumatica. Het onderwerp van dit proefschrift is de vraag of de systematische beoordeling van de

ontstekingsactiviteit en andere ziektegevolgen een waarde heeft in de behandeling van RA patiënten door de reumatoloog.

Voor de beoordeling van de ontstekingsactiviteit bij RA is het lastig dat de activiteit van het ontstekingsproces in de praktijk niet zeer nauwkeurig gemeten kan worden. Het is echter wel mogelijk om de mate van ontstekingsactiviteit af te leiden. De ziekteactiviteitsscore (DAS) wordt berekend uit het aantal pijnlijke en gezwollen gewrichten, de bloedbezinkingsnelheid en het algemeen welbevinden; een lage DAS duidt op een lage ontstekingsactiviteit. De DAS kan daarom gebruikt worden als maat om de hoeveelheid anti-reumatica en andere ontstekingsremmers af te stemmen (**Hoofdstuk 2**). Als de DAS regelmatig gemeten wordt, geeft dit ook een goed beeld van het verloop van de ziekte.

Er is al enige ervaring opgedaan met het gebruik van een gecomputeriseerd evaluatiesysteem waarmee reumatologen het ziekteverloop van RA patiënten kunnen volgen (**Hoofdstuk 3**). Het evaluatiesysteem maakt niet alleen gebruik van de DAS, maar ook van patiëntenvragenlijsten over ontstekingsactiviteit (RADAI) en functiebeperkingen (HAQ) en metingen van de ontstane gewrichtsschade (röntgenfoto's). De reumatoloog en de patiënt kunnen nu in maat en getal zien hoe hoog de ontstekingsactiviteit nu is en hoe hoog zij geweest is. Bovendien kunnen ze als behandeldoel een bepaalde mate van lage ontstekingsactiviteit afspreken en na enige tijd controleren of dat gehaald is. Een gecomputeriseerd evaluatiesysteem is daarmee een hulpmiddel voor het nemen van beslissingen over medicijngebruik, wat niet wil zeggen dat het gebruikt kan worden als "kookboek".

De RADAI patiëntenvragenlijst heeft als doel om op een relatief eenvoudige manier de ontstekingsactiviteit te meten zonder dat daar een arts voor nodig is, wat bij de DAS wel het geval is. De RADAI werd getest in een onderzoek met 584 RA patiënten (**Hoofdstuk 4**). De RADAI bleek inderdaad iets te zeggen over de mate van ontstekingsactiviteit: als de RADAI laag is, is de DAS ook laag, als de RADAI hoog is, is de DAS ook hoog. Deze samenhang van de RADAI bestaat ook met een aantal andere belangrijke maten, als: opinie van de arts, het aantal pijnlijke gewrichten en de vragenlijst over functiebeperkingen. De RADAI en de DAS zijn echter niet dusdanig identiek dat ze elkaar kunnen vervangen.

Net als de DAS zou de RADAI vragenlijst geschikt moeten zijn om het verloop van de ontstekingsactiviteit bij RA patiënten te volgen. Om daar achter te komen werd de gevoeligheid voor veranderingen van de RADAI getest (**Hoofdstuk 5**). Er werd daarbij gebruik gemaakt van de gegevens van een eerdere studie, waarin veel RA patiënten een acute toename van ontstekingsactiviteit hadden gekregen, ondanks de medicijnen. De RADAI en de DAS bleken naderhand beide even goed te zijn geweest in het registreren van opvlammende ontstekingsactiviteit.

Theoretisch is het gebruik van een gecomputeriseerd evaluatiesysteem in de behandeling van RA patiënten zinvol, het is echter lastig om dat ook te bewijzen.

Van 264 patiënten werd gedurende een controleperiode van 3 maanden de ontstekingsactiviteit gemeten met de RADAI, zonder dat de reumatologen de uitslagen kenden (**Hoofdstuk 6**). Na deze 3 maanden kregen de reumatologen 12 maanden de beschikking over een gecomputeriseerd evaluatiesysteem. Het evaluatiesysteem bleek voor de helft van de patiënten ook daadwerkelijk gebruikt te worden, maar dat hing niet samen met de hoogte van de ontstekingsactiviteit. Bij patiënten met een hoge ontstekingsactiviteit daalde de RADAI het meest, vooral als de reumatoloog het evaluatiesysteem gebruikt had. Toch bleef bij veel patiënten de medicatie onveranderd. Om bij een groter aantal RA patiënten de ontstekingsactiviteit te doen dalen, lijkt het daarom noodzakelijk om behandelingsrichtlijnen toe te voegen aan een evaluatiesysteem.

Om een indruk te krijgen of behandelingsrichtlijnen een invloed hebben op de dosis van de voorgeschreven anti-reumatica en daarmee op de ontstekingsactiviteit van RA patiënten, is gekeken naar een eerdere studie die gebruik maakte van zowel behandelrichtlijnen als de DAS (**Hoofdstuk 7**). Het leek dat het gebruik van doses die hoger waren dan de richtlijnen voorschreven, niet extra veel effect hadden. Maar het gebruik van doses die lager waren dan de richtlijnen voorschreven leek nadelig te zijn voor het behandelresultaat, terwijl er niet minder bijwerkingen veroorzaakt leken te worden. Het zou daarom best mogelijk kunnen zijn dat het gebruik van de DAS en van richtlijnen voor het bepalen van de dosis van anti-reumatica in de behandeling van RA patiënten een voordeel oplevert.

Om de redenen van gebruik en onbruik van een gecomputeriseerd evaluatiesysteem te achterhalen zijn 105 reumatologen die zo'n systeem wel eens gebruikt hebben toevallig

uitgekozen en benaderd met een vragenlijst (**Hoofdstuk 8**). In het algemeen werd bevonden dat het evaluatiesysteem aan zijn doelstelling voldoet, maar dat om diverse redenen het systeem niet erg gebruiksvriendelijk is. De reumatologen die aangaven het systeem inderdaad vanwege de patiëntenevaluatie te gebruiken, in plaats van “voor de wetenschap” of “omdat het bij ons verplicht is”, waren meer tevreden met het evaluatiesysteem en maakten er ook meer daadwerkelijk gebruik van. Wellicht kan gerichte beïnvloeding van beweegredenen en het verhogen van de gebruiksvriendelijkheid het gebruik van een evaluatiesysteem door reumatologen vergroten.

De effectiviteit van een evaluatiesysteem in de behandeling van RA patiënten is totnogtoe onbewezen. In een nieuwe vergelijkende studie zou gekeken kunnen worden of bij de behandeling van RA patiënten het gebruik van een evaluatiesysteem met behandelingsrichtlijnen een voordeel oplevert tegenover de gebruikelijke manier van behandelen (**Hoofdstuk 9**). Daarbij is het uitgangspunt dat een evaluatiesysteem zich primair richt op de reumatoloog en niet op de patiënt. Daarom worden niet de patiënten per toeval verdeeld over de twee behandelingsmogelijkheden (“evaluatiesysteem” of “gebruikelijk”), maar zijn het dit keer de reumatologen die per toeval verdeeld worden. Alle patiënten van een bepaalde reumatoloog krijgen dus automatisch dezelfde soort behandeling. Na verloop van tijd kan er gekeken worden of in de “evaluatiesysteem” groep het aantal patiënten met een lage ontstekingsactiviteit groter is dan in de groep zonder evaluatiesysteem. Omdat in dit geval de patiënten van een reumatoloog allemaal dezelfde soort behandeling krijgen, lijken deze patiënten meer op elkaar en dragen daardoor per persoon minder informatie bij. Er zijn dan ook ongeveer 2.5 maal meer patiënten nodig dan anders. Ook in de statistische berekeningen van het uiteindelijke studieresultaat moet met de gelijkenis van de patiënten van dezelfde reumatoloog rekening gehouden worden.

Publications

Current use of health status instruments in randomised controlled trials on patients with rheumatoid arthritis. Fransen J, Stucki G. *Dis Manage Health Outcomes* 1998;3(6): 271-277.

An alternative way to display SF-36 results for individual patients. Steiner WA, Fransen J, Stucki G, Aeschlimann AG. *Phys Ther* 1998;78(10): 1118-9.

Feasibility and validity of the RADAI, a self-administered Rheumatoid Arthritis Disease Activity Index. Fransen J, Langenegger T, Michel BA, Stucki G. *Rheumatology-Oxford* 2000;39:321-7.

Clinical quality management in Rheumatoid Arthritis: putting theory into practice. Uitz E, Fransen J, Langenegger T, Stucki G. *Rheumatology-Oxford* 2000;39(5):542-9.

Responsiveness of the self-assessed rheumatoid arthritis disease activity index to a flare of disease activity. Fransen J, Häuselmann H, Michel B, Caravatti M, Stucki G. *Arthritis Rheum* 2001;44(1):53-60.

Identification of case complexity and increased health care utilization in patients with rheumatoid arthritis. Koch N, Stiefel F, De Jonge P, Fransen J, Chamot AM, Gerster JC, Huyse F, So AK. *Arthritis Rheum* 2001;45(3):216-21.

Klinisches Qualitätsmanagement bei der Rheumatoiden Arthritis. Langenegger T, Fransen J, Forster A, Seitz M, Michel BA. *Z Rheumatol* 2001;60(5):333-41.

The ICIDH-2 as a framework for the assessment of functioning and disability in rheumatoid arthritis. Fransen J, Uebelhart D, Stucki G, Langenegger T, Seitz M, Michel BA. *Ann Rheum Dis* 2002;61(3):225-31.

The merits of monitoring: Should we follow all our RA patients in daily clinical practice? Fransen J, Stucki G, Van Riel PLCM. *Rheumatology-Oxford* 2002;41(6):601-4.

Reproducibility of isometric muscle strength measurements with a hand-held pull-gauge dynamometer in cancer patients. Knols RH, Stappaerts KH, Fransen J, Uebelhart D, Aufdemkampe G. *Support Care Cancer* 2002;10:430-8.

The effectiveness of a measurement-feedback system on outcome in Rheumatoid Arthritis: a controlled clinical trial. Fransen J, Stucki G, Twisk J, Chamot A-M, Gerster J-C, Langenegger T, Seitz M, Michel BA. *Accepted by Ann Rheum Dis*

Rheumatologists' opinion on the feasibility of a measurement-feedback system in RA – and the influence of motivation. Fransen J, Daneel S, Langenegger T, Michel BA. *Accepted by Rheumatology-Oxford*

Performance of the SF-36 in Patients with Rheumatoid Arthritis: Data Quality, Scaling Assumptions and Reliability. Fransen J, Steiner WA, Stucki G, Langenegger T, Seitz M, Michel BA. *Submitted*

Influence of guideline adherence on outcome in a randomised controlled trial on the efficacy of methotrexate and folate supplementation in rheumatoid arthritis. Fransen J, Laan RFJM, Van der Laar MAFJ, Huizinga TWJ, Van Riel PLCM. *Submitted*

Design and analysis of a RCT on the effectiveness of clinical decision support on physician performance and health outcome in the management of RA. Fransen J, Twisk J, Creemers MCW, Van Riel PLCM. *Submitted*

The self-assessed Rheumatoid Arthritis Disease Activity Index: reliability and responsiveness to infliximab-induced improvements. Fransen J, Forster A, Stucki G, Michel BA. *Submitted*

Short-term and long-term effects of patient education in RA patients: a systematic review. Niedermann K, Fransen J, Knols R, Uebelhart D. *Submitted*

Abstracts

The Rehab-Cycle, a measurement-improvement approach in physical therapy. Fransen J, Steiner W, Huber E, Stucki G. *Rheumatol Europe*. 1998;27(suppl 2):179.

The RADAI, a measure of disease activity in rheumatoid arthritis. Fransen J, Häuselmann HJ, Stucki G. *Rheumatol Europe*. 1998;27(suppl 2):70.

Caveats in the assessment of reliability: the example of a walking test. Fransen J, Huber E, Stucki G. *Proceedings of the 13th congress of the WCPT, Yokohama, 1999*. (p 269).

Soft tissue mobilization in the treatment of burn scars. Beintema A, Fransen J. *European Burns Association 8th Congress, Marathon, 1999*. (p 149).

Correlates of disability in Rheumatoid Arthritis. Fransen J, Langenegger T, Stucki G, Michel BA, Uebelhart D. *12th European Congress of Physical Medicine and Rehabilitation, Antalya 2000*. (p 115).

Performance of the MOS Short-Form (SF-36) Health Survey in Swiss patients with Rheumatoid Arthritis. Fransen J, Steiner WA, Langenegger T, Michel BA, Stucki G. *Ann Rheum Dis*. 2000;59(Suppl 1):276.

Effectiveness of an ambulatory work-hardening rehabilitation program. Meyer K, Klipstein A, Huwiler HJ, Fransen J, Widmer M, Michel B, Uebelhart D. *Ann Rheum Dis*. 2000;59(Suppl 1):276.

Follow-up of treatment of rheumatoid arthritis with Infliximab (Remicade) using a quality management program. Holländer R, Forster A, Fransen J, Langenegger T, Michel BA, Hasler P. *Proceedings of the Annual Meeting of the Swiss Society for Allergology and Immunology (SSAI) 2001*:P5.

A longitudinal evaluation of the effect of a decision support system on outcome in rheumatoid arthritis. Fransen J, Stucki G, Langenegger T, Seitz M, Michel BA. *Ann Rheum Dis*. 2001;60(Suppl 1):273.

Reliability and responsiveness of the RADAI, a self-assessed rheumatoid arthritis disease activity index. Fransen J, Forster A, Uebelhart D, Michel BA. *Ann Rheum Dis*. 2001;60(Suppl 1):345.

Training, eine effektive Behandlung bei Rheumatoider Arthritis. Fransen J. *Jahreskongress der Schweizerische Gesellschaften für Rheumatologie und Physikalische Medizin und Rehabilitation, Aarau 2001*.

Measurement of health-related quality of life in low back pain patients. Abt U, Uebelhart D, Rehm J, Fransen J. *Proceedings of the 1st day of clinical research, University Hospital Zurich 2002*, [150].

The effectiveness of a measurement-feedback system on outcome in Rheumatoid Arthritis: a controlled clinical trial. Fransen J, Stucki G, Twisk J, Chamot AM, Gerster JC, Langenegger T, Seitz M, Michel BA. *Proceedings of the 1st day of clinical research, University Hospital Zurich 2002*, [152].

Reliability and responsiveness of the RADAI, a self-assessed Rheumatoid Arthritis Disease Activity Index. Fransen J, Forster A, Michel BA. *Proceedings of the 1st day of clinical research, University Hospital Zurich 2002*, [153].

Accuracy of assessing physical job demands of out-patients with low back pain, using a self-administered questionnaire. Huwiler HJ, Läubli T, Fransen J, Klipstein A, Uebelhart D. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [154].

Isometric strength measurements for muscle weakness in cancer patients. Knols R, Stappaerts KH, Fransen J, Uebelhart D, Aufdemkampe G. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [156].

Design and analysis of a randomised controlled trial on the effect of work rehabilitation on return to work. Meyer K, Fransen J, Huwiler HJ, Heitz C, Szucs T, Uebelhart D, Klipstein A. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [158].

Effects of outcome quality measurement on participating physiotherapists. Niedermann K, Steiner W, Fransen J, Uebelhart D. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [159].

Long term effects of patient education for rheumatoid arthritis patients: a systematic review. Niedermann K, Fransen J, Knols R, Uebelhart D. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [160].

Effects of a preoperative exercise program on musculoskeletal rehabilitation after lung transplantation. Niedermann K, Fransen J, Böhler A, Stucki G, Uebelhart D. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [161].

Reliability of a postural balance measurement device. Swanenburg J, Knols R, Fransen J, Uebelhart D. Proceedings of the 1st day of clinical research, University Hospital Zurich 2002, [166].

Effects of exercise and nutrition on balance and risk of falling in elderly people with decreased bone mineral density: a pilot study. Swanenburg J, Fransen J, Stauffacher M, Baschung P, Uebelhart D. 13th European Congress of Physical Medicine and Rehabilitation, Brighton 2002.

Reproducibility of isometric muscle strength measurements in cancer patients. Knols RH, Stappaerts KH, Fransen J, Uebelhart D, Aufdemkampe G. 13th European Congress of Physical Medicine and Rehabilitation, Brighton 2002.

Aiming at remission when treating Rheumatoid Arthritis with Infliximab in daily clinical practice: "Titration of disease activity". Forster A, Fransen J, Langenegger T, Seitz M, Michel BA. Ann Rheum Dis 2002;61(Suppl 1):41.

Long term effects of patient education for rheumatoid arthritis patients: a systematic review. Niedermann K, Fransen J, Knols R, Uebelhart D. Ann Rheum Dis 2002;61(Suppl 1):495.

Effectiveness of an ambulatory work rehabilitation program. Meyer K, klipstein A, Huwiler HJ, Fransen J, Widmer M, Michel B, Uebelhart D. Ann Rheum Dis. 2002;61(Suppl 1):485.

Dankwoord

Allereerst gaat mijn grote dank uit naar mijn promotor, professor Piet van Riel, voor zijn waardevolle ondersteuning, zijn zinvolle kritiek, en zijn bereidheid dit internationale promotietraject vanuit Nijmegen te begeleiden.

Grote dank ben ik ook verschuldigd aan professor Gerold Stucki en aan professor Beat Michel, die het Zwitserse deel van het onderzoek mogelijk gemaakt hebben, voor hun persoonlijke stimulering en ondersteuning door de jaren heen.

Bijzondere dank geldt ook mijn beide voormalige chefs Omega Huber en Dr. Daniel Uebelhart, die het mogelijk hebben gemaakt dat de onderzoekscultuur aan het „Institut für Physikalische Medizin” tot bloei is gekomen en die mij zoveel vrijheid en waardering hebben geschonken.

Verder wil ik graag alle mensen bedanken die, op wat voor manier dan ook, hebben bijgedragen aan de totstandkoming van dit proefschrift.

In Nederland: Anke van Gestel voor alle hulp en discussies per e-mail inzake de DAS, Dr. Jos Twisk voor zijn interesse en advies inzake MLA en GEE, Dr. Roland Laan voor het ter beschikking stellen van de data voor de “guidelines” studie.

In Zwitserland: Leokadia Zurek, Synove Daneel, Beatrice Diehl, Erika Guler, Dr. Adrian Forster en Dr. Thomas Langenegger van het SCQM team, Leanne Pobjoy voor alle editing en manuscriptcorrecties. Bovendien: Dr. Werner Steiner, Karin Niedermann, Jaap Swanenburg en de research unit, Kathrin Meyer, Alex Gossmann en alle anderen van het Ergonomie team, de ASVZ Laufgruppe, Inge-Marie Velstra en het Schoorl ski-team.

Ruud en Boet bedankt voor jullie vriendschap en de herenavonden in de “Philosophie”, ze begonnen meestal wetenschappelijk en eindigden vinologisch.

Mascha, Jurjen en Willemijn.

Curriculum Vitae

De schrijver van dit proefschrift werd geboren in 1968 te Assen. Na het doorlopen van het middelbaar onderwijs in Arnhem, startte hij in 1986 met de opleiding fysiotherapie, aan de toenmalige Arnhemse Academie voor Fysiotherapie. In 1990 studeerde hij af.

Aansluitend begon hij met de studie Gezondheidswetenschappen aan de Katholieke Universiteit Nijmegen (KUN), waar hij in 1993 afstudeerde.

Begin 1994 volgde een aanstelling als fysiotherapeut, later als hoofd van de fysiotherapie, bij het medisch centrum van Dr. B. Terrier, reumatoloog in Baden, Zwitserland. In 1997 volgde een aanstelling als fysiotherapeut en wetenschappelijk medewerker, aan de 'Rheumaklinik und Institut für Physikalische Medizin' van het universiteitsziekenhuis in Zürich, Zwitserland (Dr. G. Stucki, Dr. D. Uebelhart, Professor B.A. Michel). Tevens was hij wetenschappelijk adviseur voor het onderzoeksfonds van het 'Schweizerischen Physiotherapeuten-Verband'.

In 1999 werd naast wetenschappelijke werkzaamheden voor het 'Institut für Physikalische Medizin' begonnen aan een project in het kader van het 'Swiss Clinical Quality Management in Rheumatoid Arthritis', waarvan de resultaten in de vorm van dit proefschrift voor u liggen. In 2000 werd de registratie als epidemioloog A bereikt.

Sinds 2002 is hij als wetenschappelijk onderzoeker verbonden aan de afdeling Reumatische Ziekten van het Universitair Medisch Centrum St. Radboud te Nijmegen (Professor P.L.C.M. van Riel).